Third Triennial ESCOM Conference

Proceedings

Editor: Alf Gabrielsson

Uppsala, Sweden
7-12 June 1997
Conference Organisers

Scientific committee
Klaus-Ernst Behne
Hochschule für Musik und Theater Hannover, Germany
Irène Deleège
Centre de Recherches Musicales de Wallonie, Liège, Belgium
Alf Gabrielsson
Department of Psychology, Uppsala University, Sweden
John A. Sloboda
Department of Psychology, Keele University, UK

Organising committee
Alf Gabrielsson (Chair)
Patrik Justlin
Erik Lindström
Guy Madison
Department of Psychology,
Uppsala University

Secretariat
SLU Conference Service, Uppsala
Conference manager: Tia Eriksson
Assistants: Maria Carlson / Susanne Bengtsson

Sponsors
The conference is sponsored by:
The Bank of Sweden Tercentenary Foundation
Uppsala University
The Municipality of Uppsala.
# Table of Contents

## Keynotes

<table>
<thead>
<tr>
<th>Author</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Krumhansl, C.</td>
<td>Musical tension: Cognitive, motional and emotional aspects</td>
</tr>
<tr>
<td>Imberty, M.</td>
<td>Can one seriously speak of narrativity in music?</td>
</tr>
<tr>
<td>Imberty, M.</td>
<td>Peut-on parler sérieusement de narrativité en musique?</td>
</tr>
<tr>
<td>Hodges, D.A.</td>
<td>Standing together under one umbrella: A multidisciplinary and interdisciplinary view of music psychology</td>
</tr>
</tbody>
</table>

## Symposia

<table>
<thead>
<tr>
<th>Symposium: Tristan cor anglais Tune.</th>
<th>Organiser: Irène Deliège</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lerdahl, F.</td>
<td>Prolongational structure and schematic form in Tristan's &quot;Alte Weise&quot;</td>
</tr>
<tr>
<td>Deliège, I.</td>
<td>Wagner, Tristan's &quot;Alte Weise&quot;: Two perceptual approaches</td>
</tr>
<tr>
<td>Nattiez, J.-J.</td>
<td>The Tristan's cor anglais solo: A semiological analysis</td>
</tr>
<tr>
<td>Imberty, M.</td>
<td>Quelques jalons pour une réflexion d'herméneutique psychanalytique à propos du solo de cor anglais du IIIe Acte de Tristan</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Belardinelli, M. Olivetti</td>
<td>Analysis and modelling of cognitive dynamics underlying musical behaviour</td>
</tr>
<tr>
<td>Di Matteo, R., Rossi-Arnaud, C. &amp; Tirozzi, B.</td>
<td>Rhythm processing and entrainment processes</td>
</tr>
</tbody>
</table>
Belardinelli, M. Olivetti, Camiz, P. & Mancia, E.
A discriminative model of binary and ternary musical rhythms by means of neural networks with threshold oscillations 82

Calabrese, I. & Belardinelli, M. Olivetti
Musical abilities in deaf children: Assessment, development and training 87

Ciardi, F. Cifariello
Retrieving long term memory traces in contemporary music listening: A composer's view 91

Camiz, P.
Is music unavoidable? 97

---

Thematic sessions

**Music Cognition: Pitch, Scales, Tonality** 103

Chin, C.S.
The development of absolute pitch 105

Moreno, M. & Descombes, V.
To have it or not to have it, absolute pitch is the question 111

Levitin, D.J.
Pitch memory and the A440 Scale 117

Agmon, E.
"Octave equivalence" versus "octave relatedness": Circle versus helix; chord versus melody 122

Clough, J. & Engebretsen, N.
Scales, sets, and interval cycles: A taxonomy 128

Rahn, J.
Recent diatonic theory and Curwen's tonic Sol-Fa method: Formal models for a kinesic-harmonic system 134

Edlund, B.
Tonal closure – Fact and/or fiction 140

**Composition, Improvisation** 145

Folkestad, G., Hargreaves, D. & Lindström, B.
A typology of compositional styles in computer-based music making 147

Blumenberg, H.
Improvisation and variation between performance and notation 153

Järvinen, T.
Tonal dynamics in jazz improvisation 157
Music Performance: Representation, Practice, and Reading

Lehmann, A.C. & Ericsson, K.A.
Expert pianists' mental representation: Evidence from successful adaptation to unexpected performance demands

Brodsky, W. & Henik, A.
Demonstrating inner hearing among musicians

Jørgensen, H.
Higher instrumental students' planning of practice

Nielsen, S. Graabraek
The micro structure of practice

Nielsen, S. Graabraek
Verbal protocol analysis and research on instrumental music practice

Ginsborg, J.
Singers learning and memorising a new song: An observational study

McNeil, A.
The significance of aural development in the performing musician

Reid, A.
The meaning of music and the understanding of teaching and learning in the instrumental lesson

Servant, I., Baccino, T. & Leleu, J-L.
Reading Beethoven: Eye movements investigation

Fumeaux, S. & Land, M.F.
The role of eye movements during music reading

Musical Psychoacoustics

Rakowski, A. & Taszycka, J.
Roughness of multi-part piano chords

Sandell, G.J. & Chronopoulos, M.
Perceptual constancy of musical instrument timbres; generalizing timbre knowledge across registers

Tekman, H. G.
Perceptual effects of melodic accents

Ethnomusicology

Cordes, I.
Observations on some correspondences between ethnic music and animal calls

Lehiste, I. & Ross, J.
Folksongs as interface between speech prosody and musical rhythm
Utegalieva, S.
The image of the foreign music in perception of the Kazakh Kyuishis

Music Cognition: Melody, Rhythm

Povel, D-J.
Processes in on-line music perception

Ash, R.L.
Click migration and segmentation in Gaelic melodies

Ahlbäck, S.
A computer-aided method of analysis of melodic segmentation in monophonic melodies

Martinez, I.C.
The transcription of pitches in tonal melodies. A comparative study

Eitan, Z.
Pitch direction and melodic expectancy: A reexamination of empirical data

Thompson, W.F. & Vernescu, R.
Implicit memory for music: The development of expectancies for nontonal melodies

Drake, C., Penel, A., Bigand, E. & Stefan, L.
Tapping in time with musical and mechanical sequences

Lindström, E.
Impact of melodic structure on emotional expression

Franek, M. & Mates, J.
Tempo modulation task

Madison, G.
Rhythm and time – the influence of time per se on the organisation of temporal patterns

Parker, O.G.
Cognitive knowledge of the basic elements of music

Music Performance: Psychological and Social Aspects

Brodsky, W.
Auditivity – A cognitive style and psychological orientation based on sensory preferences for audition

Brodsky, W. & Sloboda, J.A.
The effects of whole-body acoustic stimulation on subjective relaxation, verbalisation, and visual imagery among professional orchestra musicians.

Davidson, J. & Good, J.M.M.
Social processes in Western art chamber music: Some aspects of the social psychology of the string quartet
<table>
<thead>
<tr>
<th>Author</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoshino, E.</td>
<td>Anxiety and motivation in musical performance</td>
<td>333</td>
</tr>
<tr>
<td>Lidman-Magnusson, B.</td>
<td>Factors influencing singing development in poor pitch singers</td>
<td>339</td>
</tr>
<tr>
<td>Sandgren, M.</td>
<td>Opera singers: Voice, Psyche and Soma</td>
<td>344</td>
</tr>
<tr>
<td></td>
<td><strong>Music Analysis: Composers, Works, Genres</strong></td>
<td>349</td>
</tr>
<tr>
<td>Pace, E. Kotzakidou</td>
<td>Emotions, attitudes, images and ideas in J.S. Bach's music of the imagination</td>
<td>351</td>
</tr>
<tr>
<td>Lado-Bordowsky, Y.A.</td>
<td>Debussy, lecteur de Paul Bourget 1883–1906</td>
<td>357</td>
</tr>
<tr>
<td>O'Loughlin, N.</td>
<td>Schoenbergs atonal pairs</td>
<td>360</td>
</tr>
<tr>
<td>Leleu, J-L.</td>
<td>Enoncé musical et mode(s) de structuration de pitch space, ou: de la relation composition/cognition dans un fragment de l'opus 28, I de Webern</td>
<td>365</td>
</tr>
<tr>
<td>Hettergott, A.</td>
<td>Nature, ethnoaesthetics-Turangalila: Messiaen's analogously implementing exotic sound material in his symphony</td>
<td>371</td>
</tr>
<tr>
<td>Eerola, T.</td>
<td>The rise and fall of the experimental style of the Beatles. The life span of stylistic periods in music</td>
<td>377</td>
</tr>
<tr>
<td>Theobald, E. L.</td>
<td>The analysis of Roman catholic church music to show the spiritual and socialization potentials of the song</td>
<td>382</td>
</tr>
<tr>
<td>Windsor, W.L.</td>
<td>Virtual and real listening environments in electroacoustic music</td>
<td>386</td>
</tr>
<tr>
<td>Mamalis, N.</td>
<td>Caractère musical de l'espace</td>
<td>392</td>
</tr>
<tr>
<td></td>
<td><strong>Musical Development</strong></td>
<td>395</td>
</tr>
<tr>
<td>Bertrand, D.</td>
<td>Development of decentration in music listening in 3 to 8 year old children</td>
<td>397</td>
</tr>
<tr>
<td>Lamont, A.</td>
<td>Quantitative and qualitative approaches to understanding children's perception of musical pitch in development</td>
<td>403</td>
</tr>
</tbody>
</table>
Chen-Hafteck, L.
Effects of the pitch relationship between text and melody in Cantonese songs on young children's singing 412

Chen-Hafteck, L.
Languages and singing styles of children: An acoustic analysis of three children's singing 418

Gudmundsdottir, H.R.
Children's ability to identify two simultaneous melodies 424

Bergman, D. & Gabrielsson, A.
Children's judgements of structural patterns and emotional expression in music 430

Burnsed, V.
The effects of expressive variation in dynamics on the musical preferences of elementary school students 435

Fyk, J. & Krystecki, J.
Sensitivity to pitch change direction in first grade children 440

Costa-Giomi, E.
The McGill piano project: Effects of piano instruction on children's cognitive abilities 446

Paananen, P.
Cognitive development and music composition in childhood 451

Littleton, D.
Cross-cultural perspectives on young children's music-making in the context of play 457

Music Performance: Measurements 463

Penel, A. & Drake, C.
Timing variations in music performance: Hierarchical segmentation organisation and rhythmic grouping 465

Appleton, L.J., Windsor, W.L. & Clarke, E.F.
Cooperation in piano duet performance 471

Dry, A. & Gabrielsson, A.
Emotional expression in guitar band performance 475

Baars, G. & Gabrielsson, A.
Emotional expression in singing: A case study 479

Baroni, M., Caterina, R., Regazzi, F. & Zanarini, G.
Emotional aspects of singing voice 484

Rapoport, E.
Emotion and expression in opera and lied singing- From sound analysis to brain processes 490
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Juslin, P.</td>
<td>How can music performers become better at communicating emotions to listeners?</td>
<td>493</td>
</tr>
<tr>
<td>Ashley, R.</td>
<td>Expressive performance in jazz, take 2: Ranges of expressive timing in jazz melodies</td>
<td>499</td>
</tr>
<tr>
<td>Gál, B.</td>
<td>Analyzing jazz rhythms: The microstructure of walking bass lines</td>
<td>504</td>
</tr>
<tr>
<td>Toiviainen, P.</td>
<td>Modelling the perception of metre with competing subharmonic oscillators</td>
<td>509</td>
</tr>
<tr>
<td>Berggren, U.</td>
<td>Modelling musical conventions</td>
<td>511</td>
</tr>
<tr>
<td>Atalay, B.V. &amp; Placek, R.</td>
<td>Machine versus human: Responding to the task of identifying eras for selected keyboard pieces</td>
<td>517</td>
</tr>
<tr>
<td>Schmidt, L., Seifert, U. &amp; Eichert, R.G.</td>
<td>Computational auditory scene analysis as modelling tool in musicological research on perceptual auditory organization</td>
<td>517</td>
</tr>
<tr>
<td>Cambouropoulos, E.</td>
<td>The role of similarity in categorisation: Music as a case study</td>
<td>521</td>
</tr>
<tr>
<td>Rognin, P-Y.</td>
<td>Toward a formal model of musical perception</td>
<td>533</td>
</tr>
<tr>
<td>Chouvel, J-M.</td>
<td>L'intelligence musicale: Analyse et processus cognitifs</td>
<td>539</td>
</tr>
<tr>
<td>Aksnes, H.</td>
<td>A cognitive approach to musical analysis: metaphorical projection in music</td>
<td>545</td>
</tr>
<tr>
<td>Godøy, R.I.</td>
<td>Chunking in music theory by imagined sound--producing actions</td>
<td>551</td>
</tr>
<tr>
<td>Klempe, H.</td>
<td>Polyphony as a challenge to cognitive modelling</td>
<td>557</td>
</tr>
<tr>
<td>Ericson, G.</td>
<td>How to assess schoolchildren's performance in dance?</td>
<td>563</td>
</tr>
<tr>
<td>Lagerlöf, I. &amp; Djerf, M.</td>
<td>Expression in dance performances</td>
<td>569</td>
</tr>
<tr>
<td>Friberg, A. &amp; Sundberg, J.</td>
<td>Comparing runners' decelerations and final ritards</td>
<td>571</td>
</tr>
</tbody>
</table>

Music and Bodily Movement, Dance

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lagerlöf, I. &amp; Djerf, M.</td>
<td>Expression in dance performances</td>
<td>577</td>
</tr>
<tr>
<td>Friberg, A. &amp; Sundberg, J.</td>
<td>Comparing runners' decelerations and final ritards</td>
<td>582</td>
</tr>
<tr>
<td>Topic</td>
<td>Authors</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>Music Cognition: General</td>
<td>Clarke, E.F.</td>
<td>587</td>
</tr>
<tr>
<td></td>
<td>From perception to critique: Direct realism and musical meaning</td>
<td>589</td>
</tr>
<tr>
<td></td>
<td>Dibben, N. &amp; Clarke, E.F.</td>
<td>594</td>
</tr>
<tr>
<td></td>
<td>The perception of associative structure in atonal music</td>
<td>599</td>
</tr>
<tr>
<td></td>
<td>Shifres, F.</td>
<td>605</td>
</tr>
<tr>
<td></td>
<td>Interpretative criterion and understanding of the textural hierarchy: A production-study about some internal representations</td>
<td>611</td>
</tr>
<tr>
<td></td>
<td>Cohen, D. &amp; Mondry, H.</td>
<td>619</td>
</tr>
<tr>
<td></td>
<td>Learned and innate (natural) schemata</td>
<td>625</td>
</tr>
<tr>
<td>Music Experience</td>
<td>Watt, R.J. &amp; Ash, R.L.</td>
<td>629</td>
</tr>
<tr>
<td></td>
<td>Experimental analysis of the perception of sequences of tones</td>
<td>635</td>
</tr>
<tr>
<td></td>
<td>Fung, C.V.</td>
<td>644</td>
</tr>
<tr>
<td></td>
<td>Effect of video presentation on Asian music perceptual dimensions</td>
<td>650</td>
</tr>
<tr>
<td></td>
<td>Bochkarev, L.L.</td>
<td>655</td>
</tr>
<tr>
<td></td>
<td>Social–perceptual mechanisms of musical experience</td>
<td>661</td>
</tr>
<tr>
<td></td>
<td>Sloboda, J.A., Lehmann, A.C. &amp; Parnutt, R.</td>
<td>669</td>
</tr>
<tr>
<td></td>
<td>Perceiving intended emotion in concert-standard performances of Chopin's Prelude No. 4 in e–minor</td>
<td>675</td>
</tr>
<tr>
<td></td>
<td>Finnäs, L.</td>
<td>681</td>
</tr>
<tr>
<td></td>
<td>Teenagers' experiences in typical &quot;aesthetic areas&quot;</td>
<td>687</td>
</tr>
<tr>
<td></td>
<td>Öblad, C.</td>
<td>693</td>
</tr>
<tr>
<td></td>
<td>Music in the car</td>
<td>701</td>
</tr>
<tr>
<td></td>
<td>Kleinen, G.</td>
<td>707</td>
</tr>
<tr>
<td></td>
<td>The metaphoric process. What does language reveal about music experience?</td>
<td>713</td>
</tr>
<tr>
<td></td>
<td>Gabrielsson, A.</td>
<td>719</td>
</tr>
<tr>
<td></td>
<td>Varieties of music experience</td>
<td>725</td>
</tr>
<tr>
<td></td>
<td>Anagnostopoulou, C.</td>
<td>731</td>
</tr>
<tr>
<td></td>
<td>Lexical cohesion in linguistic and musical discourse</td>
<td>737</td>
</tr>
<tr>
<td>Music and Neuropsychology</td>
<td>Tervaniemi, M., Winkler, I. &amp; Näätänen, R.</td>
<td>743</td>
</tr>
<tr>
<td></td>
<td>Timbre perception and sound duration: An ERP study</td>
<td>749</td>
</tr>
<tr>
<td></td>
<td>Gruhn, W., Altenmüller, E. &amp; Parlitz, D.</td>
<td>755</td>
</tr>
<tr>
<td></td>
<td>Neural representations of music evoked by verbally musically based learning strategies. An EEG study on cortical auditory activation patterns.</td>
<td>761</td>
</tr>
</tbody>
</table>
Altenmüller, E., Schürmann, K., Parlitz, D. & Behne, K–E.
Brain activation patterns during music processing reveal emotion–related hemispheric differences 675

Levitin, D.J. & Bellugi, U.
Musical abilities in individuals with Williams syndrome 681

Abrams, R.M., Griffiths, S.K., Huang, X., Sain, J., Langford, G. & Gerhardt, K.J.
The fetal musical environment 684

Rauscher, F.H., Robinson, K.D. & Jens, J.J.
Spatial performance as a function of early music exposure in rats (Rattus norvegicus) 688

Music Performance: Motor Processes 695

Edlund, B.
Motor patterns and musical structure 697

Pamcutt, R., Sloboda, J.A. & Clarke, E.F.
Interdependence of right and left hands in sight–read, written, and rehearsed fingerings of piano music in parallel octaves 702

Kruger, M., Lammers, M., Fuller, R., Allyn, D.A. & Stoner, L.J
Biomechanics of music performance: Moving the trombone slide 706

Music Performance: Modelling 711

Langner, J.
Multidimensional dynamic shaping 713

Friberg, A., Frydén, L. & Sundberg, J.
A rule for automatic musical punctuation of melodies 719

Tangian, A.
On the technology of interpretation 724
Keynotes
The metaphor of tension is frequently applied to music. My talk will consider this metaphor from three different points of view: cognitive, motional, and emotional. These points of view are based on Alf Gabrielsson's (1973a, b) classic study of rhythm qualities. This extensive series of experiments on the experience of rhythm uncovered three classes of dimensions. The first class includes cognitive-structural dimensions, such as meter (whether duple, triple, and so on), accent on the first beat, clearness of accent, uniformity vs. variation, and simplicity vs. complexity. The second class includes movement-motion dimensions, including rapidity, rate of elements, forward motion, and such movement associations as dancing vs. walking, rocking vs. knocking, graceful vs. thumping, and solemn vs. swinging. The third class includes emotion dimensions such as vital vs. dull, excited vs. calm, rigid vs. flexible, and solemn vs. playful. Thus, rhythms appear to have effects on three psychological levels: cognitive, motoric, and affective.

Descriptions of music, particularly in the Western tonal-harmonic style, often refer to tension and relaxation. According to Ian Bent (1987), this way of describing music arises in early 20th century musical analysis and was influenced by Gestalt psychology. Three Gestalt principles were extended to music: closure, which automatically completes partially incomplete patterns; the phi phenomenon, which interpolates a link between two separate occurrences; and Prägnanz, which seeks the simplest possible perceptual organization. In addition, figure-ground organization motivated the analysis of music by reductions to show the essential underlying structure. Musical form was considered a type of whole or Gestalt. Phrases, motives, rhythms, and other musical patterns bear specific relationships with one another, and together determine the musical form. The musical form, in turn, exerts influences on the perception of the musical components, just as a visual form influences the perception of its components. Thus, the metaphor of tension provides a natural link between musical structure and perceptual and cognitive responses to music.

Theorists often emphasize the dynamic nature of musical tension, with continuous variations over time. For example, Leonard Meyer (1967, p. 43) states "... music is a dynamic process. Understanding and enjoyment depend upon the perception of and response to attributes such as tension and repose, instability and stability, and ambiguity and clarity." The dissertation by Frede Nielsen (1983) demonstrated that this aspect of music perception can be brought under experimental control. In the experiments, listeners pressed a pair of tongs together to indicate the experienced degree of tension. The amount of pressure applied was recorded continuously while music by Haydn and R. Strauss was played. This method produced strikingly regular tension curves, with smaller waves of tension superimposed on larger waves of tension. These judgments could be related to specific musical factors, including dynamics, timbre, melodic contour, harmony, tonality, and repetition. According to Nielsen, "In the musical structure of strata, 'tension' is assumed to be placed in the middle region of the object, connecting structural characteristics of the surface level with more deeply located strata of emotion and other strata of meaning." (p. 316)

Krumhansl (1996) took a similar approach using a computer interface to study dynamic changes in tension in the first movement of Mozart's Piano Sonata in Eb major, K. 282. The talk will review that experiment with special emphasis given to predictions of Lerdahl's (1988) pitch-space model. The model quantifies three pitch-space distances. Pitch-space i distance is the distance between two chords on the diatonic level. Pitch-space j is the distance between the chords on the cycle of fifths for chords. Pitch-space k is the number of tones that are not shared by the two chords counted at all levels of the pitch-space model. Most importantly, inherited value is the sum of the pitch-space i, j, and k values for all events superordinate to an
event in the prolongational tree. Thus, this value takes into account the hierarchical embedding of events in the prolongational structure. To these four values, Lerdahl (1996) adds three measures of surface dissonance. Scale degree codes whether the melodic tone is contained in the supporting triad. Inversion codes whether the triad is in root position or inversion. Nonharmonic tone codes for the presence of tones not in the chord. This gives seven quantitative predictors to compare with the data.

The quantitative predictors modeled the tension judgments well, with the pitch space and inherited value being the strongest of the seven variables in the model. A subsequent analysis using a moving average of the predictor variables provided an even better fit of the data, suggesting that listeners integrate the musical information over time. Tension was also associated with pitch height, dynamics, mode, denial of expected patterns within the piece, interruption of ongoing harmonic processes, and other attributes of the music. Thus, dynamic judgments of musical tension can be attributed quite precisely to theoretical analyses of musical structure and related to attributes that have received considerable attention in the literature on music cognition.

In addition, this experiment found that the tasks had rather consistent relationships with one another and with performed tempo. These are of interest in comparison with discourse structure as studied by Wallace Chafe (1994). Judgments that a section end had occurred were soon followed by judgments that a new musical idea had been introduced (and, conversely, judgments that a new musical idea had occurred were almost always preceded by judgments that a section end had occurred). This is analogous with the placement of new topics at the beginning of discourse units. Section-end responses corresponded to peaks in tension followed by rapid decreases in tension and slower tempos. This may be analogous to the patterns of phrase final lengthening and pauses at the end of discourse units. In contrast, new musical ideas were introduced when tension was at a low level and the tempo was neutral. This may correspond to what Chafe (1994) identifies as starting points or points of departure that are prepared by the larger linguistic context. Finally, an asymmetric pattern of tension within segments found in this experiment, possibly corresponding with how ideas are developed and completed within linguistic units.

The next part of the talk examined possible parallels between musical tension and bodily movement. The relationship between music and bodily movement has been approached from a variety of directions. For example, systems have been proposed for representing music by bodily movement. These proposals are based on the premise that the two modalities can potentially communicate similar expressive qualities. In support of this, a study by Robert Francêz and Bruchon-Schweitzer (1982) found actors could successfully convey musical expression in videotaped recordings of their movements. Johan Sundberg and collaborators have proposed two specific relationships between music and bodily movement: between musical tension and vocal effort (Sundberg, 1987), and between timing patterns in musical rubato and the decrease of footstep rate of a runner coming to a stop (Kronman & Sundberg, 1987; Sundberg, Friberg, & Fryden, 1991). Jane Davidson (1993) reported perceptual experiments investigating whether auditory and visual modalities convey similar information about expressiveness in musical performances. Observers reliably perceived the degree of expressiveness whether they only heard the music, or only saw point-light displays of the performers' motions, or both heard the music and saw the visual display.

Dance is another domain in which to study the relationships between music and bodily movement. A number of considerations suggest the potential for rather close mappings between music and dance. Both consist of temporally extended sequences that subdivide into structured units (e.g., Lasher, 1981; Knuthansl, 1996). The beginnings of the units are marked by stable, preparatory elements, followed by unstable elements that resolve with pattern completion. In one of the few available psychological studies of dance, Margo Lasher (1981) demonstrated the existence of a cognitive schema for such units. Observers recognized preparatory–completing segments more accurately than completing–preparatory segments.
Moreover, they falsely recognized preparatory–completing segments that had not in fact appeared in the dance sequence. These results parallel those showing that unstable elements in music are poorly recognized and are frequently confused with more stable elements, specifically for tones, harmonies, melodic sequences, rhythm, and meter. In addition, parallels between music and dance (choreomusical parallels) may be found in tempo, dynamics, texture, contour, and the structuring of larger–scale hierarchically organized formal units.

This exploratory study (Krumhansl & Schenck, 1997) examined the extent to which observers are sensitive to mappings between music and dance when correspondences exist. For this purpose, we selected a piece that intuitively appeared to contain many choreomusical parallels, a dance choreographed by George Balanchine to Mozart's Divertimento No. 15 in Bb major (KV 287). In the experiment, the participants either heard the music (Music Only condition), saw the dance (Dance Only condition), or heard the music and saw the dance (Both Music and Dance condition). They performed four on–line tasks on successive presentations of the stimulus materials. Three of the tasks were used in the previous study (Krumhansl, 1996): judgments of section ends, amount of tension, and new ideas. The fourth task, the task of judging the amount of emotion expressed, is patterned after Davidson (1993), except that it is judged continuously during the piece. Finally, at the end of the experiment participants judged the emotional qualities so that we could determine the degree of similarity in the emotion expressed by the stimulus materials in the different conditions.

Of primary interest are the comparisons across the three conditions (Music Only, Dance Only, and Both Music and Dance). If choreomusical parallels are psychologically real, then the data from the Music Only and the Dance Only conditions should correlate even though they contain none of the same stimulus materials. The music and dance can then be analyzed to determine the elements that induce similar cognitive and affective representations of this piece. The experiment also examined the combined effect of the music and dance. Specifically, the data from the Both Music and Dance condition can be tested to see whether it can be predicted by the Music Only and Dance Only conditions. If so, then the relative weights can be assessed, as well as possible interactions between the two components. Also of interest is whether the two components evoke similar affective responses, especially for the Music Only and Dance Only conditions which contained none of the same stimulus materials. Of secondary interest are the temporal patterns that obtain among the four tasks. The previous study (Krumhansl, 1996) found consistent relationships among judgments of section ends, tension, and new ideas. Do these relationships also apply to dance?

The results of the experiment quite strongly supported the idea that observers are sensitive to correspondences between music and dance. Analyses of the data showed correlations between the Music Only and the Dance Only conditions in all four on–line tasks. That is, section ends and new ideas tended to coincide in the music and the dance, and the curves of tension and emotion expressed were also similar. This was true even though these conditions used completely different stimulus materials. Some minor differences between conditions should be noted, however. The delineation of temporal units by section ends and new ideas was clearest in the Music Only condition, which also exhibited the greatest and most regular variability in tension and emotion expressed. In contrast, the segmentation in the Dance Only condition appeared primarily at a higher hierarchical level, with fewer subdivisions, and the profiles of tension and emotion expressed were somewhat smoother, showing less local variation. In addition, unlike the Music Only condition, these profiles exhibited a general trend toward a peak in the final section of the piece, replicating on the highest level the pattern found in subsections.

An advantage of the on–line psychological measurements in this study is that they can guide the analysis of the materials for elements that correspond in the music and dance. As noted above, temporal segmentation in the dance occurred primarily at higher hierarchical levels. On this level, the dance consists of three main sections, each corresponding to the three sections of the minuet. The first section, or first minuet, is performed by the entire corps de
ballet; this is followed by the duet pairings in the trio section, a marked change of style and form in both the dance and the music signaled by the regrouping of the dancers into a tableau formation and a change of key to G-minor; the return to the A section at the end is mirrored by a return to the use of the full complement of eight dancers, who perform some of the material from the earlier performance of A, but also execute variations on many of the steps performed by the duet pairs in the trio. Consequently, the dance reflects the essence of the minuet form without resorting to a 1:1 imitation of its structural material. Each of these dance units subdivides into units that correspond with smaller musical sections. In the first and third sections, each of these subsections is marked by regrouping of the dancers into a new formation. Structural parallels are also evident at smaller subdivisions. For example, the first six measures contain three repetitions of a transposed ascending melodic motive that corresponds with the two lines of dancers moving in opposition three times.

It was possible to find quite a few other intrinsic parallels in this selection. Rhythmic parallels are easily observed, as Balanchine keeps to a strict 3/4 pulse, and rarely deviates from the two-, four-, and eight-measure hierarchies of the music. Overt dynamic parallels also exist, as fortes and pianos are accompanied by corresponding larger, stronger movements, or smaller and more delicate movements. For example, at a piano dynamic the dancers execute delicate pointe work and graceful arabesques; as the dynamic changes suddenly from piano to a solid forte, the dancers cross through each other with forceful jetés, stronger arm positions, and emphatic turns. Similar correspondences between the highness of the notes in the passage is mirrored in the higher steps and leaps of the dancers. Other qualitative choreomusical parallels are correspondences between the "sharpness or smoothness" of the dance steps and arm positions and the use of staccato or legato markings in the score. The legato, arc motive was accompanied by smooth, graceful, "low-key" movements, while the triumphant, marcato horn arpeggio is fleshed out by a similarly marcato pas de chat jump. These kinds of features can account for the strong correspondences between the music and the dance reflected in the four on-line tasks. Thus, there appear to be a large number of dimensions available in the two modalities to establish structural mappings between the two domains, suggesting non-accidental mappings between music and the bodily movement in the dance.

An additional question of interest in the present study was how music and dance combine in the psychological representation. To address this, the results from the Both Music and Dance condition were compared with the Music Only and the Dance Only conditions. One possibility is that the results in the former are a combination of the other two conditions that combine in an additive, non-interactive fashion. Another possibility is that there is an interaction between the two, in which the combined effect is greater than or different from the sum of the two components. The analyses of the present data support the former possibility. The data in the Both Music and Dance condition were modeled well by a weighted additive combination of the data in the Music Only and the Dance Only conditions. For all tasks, the analyses found somewhat greater weight for the Music Only data than the Dance Only data. It is noteworthy that this non-interactive relationship between music and dance is similar to that found in some perceptual and cognitive-neuropsychological studies for musical pitch and rhythm.

It is more difficult to account for the similar judgments of global emotions found in the Music Only and Dance Only conditions. Both conditions produced high ratings of Amused, Contented, Happy, and Interested. Given the present data, we can only speculate on what produces these similar emotional responses. Some factors that seem plausible for this particular selection are the tempo of the music and the dance, melodic and choreographic gestures, and the high degree of symmetry and regularity in both music and dance. Additional factors may be the pacing of segment endings and new ideas, and the experienced contours of tension and emotion expressed. The present finding that music and dance can produce similar emotional representations suggest they are also carried to some extent by properties with cross-modal analogs. Related studies on music and film suggest that music plays an important
function in determining the affect and mood of the film and directing attention and adding meaning to particular events. Such cross-modal studies promise to contribute to our understanding of how music acquires its semantic content and affective qualities by examining the ways it coordinates with these other art forms.

Finally, in all three conditions the four on-line tasks showed consistent temporal patterns with one another. New ideas were judged as occurring at section beginnings when levels of tension and emotion expressed were low. These levels tended to increase throughout the sections, reaching peaks just before the section ends, and then declining rapidly. The finding that this temporal pattern appeared in all conditions suggests that a general schema for temporal organization is operating. This pattern was also found in the previous study by Krumhansl (1996), where it was compared with discourse structure in language (Chafe, 1994). Again, it should be noted that the generality of the schema may be limited to only certain styles of music and dance. Both the present and the previous studies used pieces by Mozart written in the classical style, which is frequently compared with discourse. That the pattern also appears in the dance may be due to Balanchine's attentiveness to the music, as noted above. Thus, while the present results suggest that this schema for temporal organization is available for both music and dance, it should not be assumed to be obligatory for all styles. Additional research is needed to explore these relationships in other styles of dance.

The final part of the talk focuses on emotional reactions to music, taking the approach of correlating psychophysiological measures with dynamic changes in emotional reactions to music. A philosophical debate about musical emotions concerns whether music produces emotional changes in listeners (the "emotivist" position) or simply expresses emotions that listeners recognize in the music (the "cognitivist" position). Peter Kivy (1990), one of the strongest proponents of the latter position, distinguishes between musical "cognitivists" and musical "emotivists" as follows: "Those I am calling musical emotivists believe that when, under normal circumstances, musical critics, theorists, or just plain listeners call a piece of music (say) "sad," it is because it makes us sad when we listen to it; and what they mean by "sad" music, I will assume, is music that normally arouses sadness in the normal listeners. The musical cognitivists, like the emotivists, believe that it is proper sometimes to describe music in emotive terms. But unlike the emotivists, they do not think that sad music is sad in virtue of arousing the emotion in the listeners. Rather, they think the sadness is an expressive property of the music which the listener recognizes in it..."

Kivy rejects the emotivist position absolutely except to admit the possibility that prior associations or psychopathology may produce musical emotions: "If a piece of music makes someone sad, or frightened, or despairing, or angry, you can be sure the reaction is either personal or pathological." (p. 170) His rejection rests on a number of arguments. Most relevant here is that he finds theories of how music produces emotions, which include conflict theories, psychoanalytic theories, behaviorist theories, and physiological theories, to be unconvincing: "[The theories]... range from the wildly false through the uselessly true to the highly controversial..."scientific" theories of the emotions come, and they go; and there seems... to be something deeply wrong with the way they are put in the service of musical aesthetics by the emotivists.... I have nothing against the scientific study of the human psyche or against scientific theories of emotion per se, and I am certainly not arguing that because we do not yet have a successful psychology and physiology of the emotions (if that is indeed true), we can never have one." (p. 149) However, "whatever turns out to be the true "scientific" theory (or one of the true theories) of human emotion, music presents it with a peculiar anomaly.... Music will present the triumphant psychology with nothing to explain more deeply or more broadly, or to explain away, about how it arouses fear and anger, despair and hope, joy and sorrow and the rest." (p. 157) In the meantime, "that there are no behavioral symptoms of listeners actually experiencing [emotions] when attending to music, is... decisive." (p. 151)

The cognitivist position is represented also by Leonard Meyer (1956, p. 8), who states: "The depiction of musical moods in conjunction with conventional melodic or harmonic
formulas, perhaps specified by the presence of a text, can become signs which designate human emotional states... And it may well be that when a listener reports that he felt this or that emotion, he is describing the emotion which he believes the passage is supposed to indicate, not anything which he himself has experienced." He dismisses the importance of physiological changes in response to music, such as heart rate and skin conductance, as being inconclusive on two grounds. First, "...no relation can be found between the character or pattern of the musical selection evoking the response and the particular physiological changes which take place. These changes appear to be completely independent of any particular style, form, medium, or general character." Second, "the listener brings to the act of perception definite beliefs in the affective power of music. Even before the first sound is heard, these beliefs activate dispositions to respond in an emotional way, bringing expectant ideo-motor sets into play. And it seems more reasonable to assume that the physiological changes observed are a response to the listener's mental state rather than to assume that tone as such, in some mysterious and unexplained way, bring these changes about directly." (p. 11)

Nonetheless, Meyer (1956, p. vii) noted the potential rewards of an increased understanding of musical emotions, "Not only does music use no linguistic signs but, on one level at least, it operates as a closed system, that is, it employs no signs or symbols referring to the non-musical world of objects, concepts, and human desires....This puzzling combination of abstractness with concrete emotional and aesthetic experience can, if understood correctly, perhaps yield useful insights into more general problems of meaning and communication..." Toward this end, he develops an account of musical emotions based on the fundamental premise: "Affect or emotion--felt is aroused when an expectation -- a tendency to respond -- activated by the musical stimulus situation, is temporarily inhibited or permanently blocked." (p. 31) Arguing from numerous musical examples, he proposed that these expectations are derived from both general psychological principles (such as Gestalt principles of perceptual organization) and knowledge of the style (such as tonality, harmonic progressions, and musical form). What is important to emphasize here is that Meyer's (1956) account ascribes musical emotions to variations in tension and release from tension operating over short spans of time, rather than to more general musical characteristics that may establish moods over longer durations.

The cognitivist position stands in stark contrast to the assumption either implicit or explicit in studies of musical emotions, that musical emotions are of the same kind felt in non musical situations. Most available studies use verbal reports of emotional responses to music, and these find considerable consistency across individuals. The results of these studies, and advances in psychological theory and methodology for studying emotions, motivate further study of musical emotions. The experiment reported here takes the approach of measuring psychophysiological changes during music listening. According to a recent survey of the field of psychophysiology (by Kenneth Hugdahl, 1995, p. 8), "Psychophysiology is the study of brain–behavior relationships in the framework of peripheral and central physiological responses." Recording of psychophysiological responses is regarded as a "window" into the brain and mind. These responses include measures of the central nervous system (through electroencephalograms, event–related potentials, and more recently brain imaging techniques) and peripheral nervous system (electrodermal activity, heart and blood circulation, respiration, and muscular activity). The present study exclusively considers measures of peripheral nervous system function.

It is of an exploratory nature. It sought to obtain a fairly wide spectrum of physiological measures of cardiac, vascular, electrodermal, and respiratory function on the same subjects while listening to music. These were taken in the laboratory of Robert Levenson. The measures were taken continuously (at one–second intervals) during the music. Six excerpts were chosen to represent (two selections each) the emotions of sad, fear, and happy. One question of interest is whether different configurations of physiological responses correspond to these different emotions. Each of the excerpts was approximately three minutes
in duration. Each excerpt was preceded by a 90-second period of silence during which baserate measures were collected. These values were subtracted from the pre-music measure to control for individual differences and overall changes that might occur during the experimental session. In addition, listeners gave self-report ratings of their emotional responses during the excerpts. However, the primary emphasis is on dynamic changes that occurred during the musical excerpts. Therefore, independent groups of listeners made dynamic ratings of the degree of sadness, fear, happiness, and tension experienced while listening.

The first finding was that the six selections did, as indicated by self-report measures, produce the intended emotions. Both groups of listeners, those who gave dynamic emotion quality ratings and those whose physiological measures were taken, reported very similar responses to the selections. Intersubject consistency was also very strong. The subjects judged their response to the excerpts by Albinoni (Adagio in G minor for Strings and Orchestra) and Barber (Adagio for Strings) to be strongest for the emotion of sad. They judged their response to the excerpt by Holst (Mars — the Bringer of War from The Planets) and Mussorgsky (Night on Bare Mountain) to be strongest for the emotions of anxious, afraid, and surprised. They judged their responses to the excerpt by Vivaldi (La Primavera — Spring — from The Four Seasons) and Hugo Alfven (Midsommarvaka) to be strongest for the emotion of happy, followed by amused and contented. That the latter two cases showed graded responses between intuitively similar emotions is consistent with the patterns of responses found, for example, by Hevner (1936), and supports the idea that musical emotions are not differentiated at fine-grained levels within basic emotion categories.

A two-dimensional representation of the emotional reactions to the six excerpts was recovered by a factor analysis of these self-reports. The circular representation, with emotions separated according to positive versus negative valence and degree of activity, corresponds with that typically found in studies of non musical emotions. The grouping of emotion terms was clear within this two-dimensional space. In it, the emotion of afraid was close to a number of other emotions including surprised, anxious, and contemptuous. The emotion of happy was close to a number of other emotions including amused, contented, and relieved. Thus, the emotional responses reported in reaction to the musical excerpts not only coincided with the intended emotions, but produced a pattern of relationships between emotions that conformed with those that arise from studies of non musical emotions.

The dynamic ratings of emotion over time also showed the intended emotions of the chosen excerpts. All six excerpts were judged, by different groups of listeners, on a second-by-second basis for the degree of sadness, fear, and happiness. As expected, the ratings of sad were highest for the Albinoni—Barber pair, the ratings of fear were highest for the Holst—Mussorgsky pair, and ratings of happy were highest for the Vivaldi—Alfven pair. In general, the ratings of the intended emotions maintained a fairly high level that was sustained throughout the excerpts with local variations. This suggests that music may be a particularly good stimulus for inducing and maintaining an emotion for extended periods of time. Dynamic ratings of tension were also obtained for possible interest in connection with other studies measuring tension. Tension correlated most strongly with fear, although it also correlated quite strongly with the dominant emotion of the particular excerpt. A strong correlation with dynamic sad ratings was found for the Albinoni and Barber excerpts, and a strong correlation with dynamic happy ratings was found for the Vivaldi and Alfven excerpts. This suggests that tension is a multivalent quality, influenced by the predominant emotional response to the music.

With these results in hand, we can now turn to the psychophysiological measures themselves. Given the complexity of the data, with twelve different physiological measures and second-by-second measures for six three-minute long excerpts, it seemed desirable to conduct analyses that would look for convergences. A number of general patterns emerged from these analyses. The first result was that the presence of music significantly affected all
twelve physiological measures, and the direction of the changes was the same for all three excerpt types. The changes were: longer cardiac interbeat interval (increased IBI, slower heart rate), longer pulse transmission times to ear and finger (increased FPTT and EPTT) and reduced blood amplitude at the finger (decreased FPA), decreases in respiration intercycle interval, depth, and respiration–sinus asynchrony (decreased ICI, which means a faster breathing rate, and decreased RD and RSA), higher systolic, diastolic, and mean arterial blood pressure (increased SBP, DBP, and MAP), and lower skin conductance and finger temperature (decreased SCL and TEM). Thus, the presence of music had effects on all of these physiological measures. Speculatively, this might be taken as support of Meyer's (1956) claim that music has a general effect on physiology because it induces an "expectant ideo–motor set" activating dispositions to respond in an emotional way. What is more important, it argues against the cognitivist position taken, for example, by Kivy (1990) and his claim that there are no behavioral indicators that listeners experience emotions when listening to music. Not only do listeners verbally report emotional responses to music with considerable consistency, music appears to produce physiological changes as well.

The central question that remains, then, is whether there are emotion–specific patterns in the psychophysiology. This question has a long and vexed history going back to William James (1890) who proposed that different emotions are associated with unique patterns of physiological changes. It was hypothesized that these patterns differentiate the emotions, in other words, identification of the emotion felt is determined by a reading of the physiological responses. An influential study by Schacter and Singer (1962) undercut this position by showing that undifferentiated arousal resulted in different reports of emotions depending on the subject's cognitive response to external events. More recently, the first position -- the physiological differentiation of emotions -- has been strengthened by a series of studies by Paul Ekman, Robert Levenson and colleagues (e. g., Ekman, Levenson, & Friesen, 1983; Levenson, Ekman, & Friesen, 1990; Levenson, 1992, 1994; see however, Zajonc and Mcintosh, 1992; Cacioppo, Klein, Berntson, and Hatfield, 1993). These studies suggest consistent differences in some physiological correlates of anger, fear, sadness, and happiness.

A number of analyses were conducted to assess emotion–specific physiology in the present study. These focused on the physiological data recorded at one–second intervals throughout the musical selections (corrected for baserate levels during the pre–music interval) because, as suggested by Meyer (1956), these dynamic measures may exhibit time–locking to structural properties of the music. The primary analysis correlated the dynamic physiology ratings with the dynamic ratings of emotion quality (sadness, fear, and happiness). Strongest correlations with ratings of sadness were found for: cardiac interbeat interval (positive correlation with IBI), the three measures of blood pressure (positive correlations with SBP, DBP, MAP), skin conductance level (negative correlation with SCL), and finger temperature (negative correlation with TEM). Strongest correlations with ratings of fear were found for: finger and ear pulse transmission time (positive correlations with FPTT and EPTT), and finger pulse amplitude (negative correlation with FPA). Strongest correlations with the ratings of happiness were found for the respiration measures of intercycle interval, respiration depth, and respiration–sinus asynchrony (negative correlations with ICI, RD, and RSA).

These patterns were substantiated and clarified somewhat in further analyses. A factor analysis showed that the patterns of physiological changes seemed to separate by system. The dynamic emotion ratings of sad, fear, and happy, and were aligned with distinct factors corresponding to different physiological systems. Sad was associated with changes in measures of cardiac and electrodermal systems, fear with changes in cardiovascular measures, and happy with respiration measures. Despite the significance of many of the correlations, it should be emphasized that their magnitudes were generally quite small. However, most of the physiological changes either increased during the musical selection or remained at a fairly constant level over time; only a few decreased toward baserate levels. These results suggest that musical emotions may indeed be reflected in psychophysiological measures, supporting
the emotivist position that musical emotions are felt emotions. The second analysis that correlated the dynamic physiology ratings with the dynamic ratings of emotion quality (sadness, fear, and happiness).

Finally, I consider correspondences between the present results and those obtained in previous studies of emotion-specific physiology. First, I will consider the summary of the data by Zajonc and McIntosh's (1992) who considered the results from the directed facial action task and the relived emotions task in the work of Paul Ekman and Robert Levenson. Most of the significant differences found in these studies were for the directed facial action task. The results of these studies were compared with the physiological averages in this study. Considering first heart rate, insofar that these studies found differences, the measure tended to higher for sad and fear than for happy. The inverse order would be predicted for the cardiac interbeat interval measure used here, namely, interbeat interval would be higher for happy than sad and fear. In contrast, the present measure was higher for sad than fear or happy. For finger temperature, the previous studies found no significant differences among sad, fear, and happy. Here, finger temperature was higher for happy than sad and fear excerpts. For skin conductance, the previous data would predict the lowest value for the happy excerpts. The present results found, instead, that the lowest value was for the sad excerpts. Similar disparities were found with the cross-cultural data in the study by Levenson, Ekman, Heider, and Friesen (1992) in which the directed facial actions task was performed by the Minangkabau of West Sumatra.

To sum, quite a few of the differences found in these studies, especially in the directed facial actions task, were opposite those found in the present experiment. Possibly, as Frans Boiten (1996) has suggested, the physiological changes in the directed facial actions task may depend more on the difficulty of producing the facial expression than the underlying emotion. More generally, emotion physiology may depend on preparation for overt responses as emphasized by Levenson (1994) and others. Music listening is often not accompanied by overt responses to the music. In this light, it may be noteworthy that the effects of music on a number of physiological measures was similar to the suppressed emotion condition in the study by James Gross and Robert Levenson (1993). In that condition, subjects were instructed not to make overt responses while watching a disgust-inducing film. The physiological measures were compared to a control condition with no such instruction. Similar differences as those found here for music obtained for cardiac interbeat interval (IBI), finger pulse amplitude (FPA), finger temperature (TEM), respiration intercycle interval (ICI), and respiration depth (RD). These similarities suggest that suppression of overt action during music listening may also affect physiological measures.

Somewhat more agreement was found with the summary of data by Cacioppo et al. (1993). Of all seven cases with significant differences in this and the reviewed studies, the direction of the difference found was the same. For heart rate, the present study agreed with Averill (1969, maximum rise, using film manipulation of emotion) and Tourangeau and Ellsworth (1979, maximum rise, also using film manipulation). For finger temperature, skin conductance level, and finger pulse amplitude, the present study agreed with Stemmler (1989, real life manipulation, where fear was induced by a scary radio play and music and unexpected darkness, and happy was induced by a nice experimenter, extra monetary bonus, and a shorter experiment). Finally, both systolic blood pressure and diastolic blood pressure changes in the present experiment agreed with those found by Averill (1969, using a film manipulation). It may be significant that all the manipulations in these studies were extended over time as were the musical excerpts in this experiment. Thus, the question of emotion-specific physiology remains somewhat elusive.

These three research projects explore the dynamic character of responses to music. They demonstrate that continuous on-line measurements can be made without interrupting or otherwise artifically manipulating musical materials. The data using these methods showed precise time-locking to musical events and considerable reliability across subjects and
repetitions. The results suggest that such methods may prove useful in linking cognitive-structural aspects of music, that have been the primary focus of recent research, to the motional and emotional aspects of music. It is hoped that these and related studies will encourage empirical research and theorizing about the connections between cognitive, motoric and emotional responses that give meaning to musical sounds.

References
Can one seriously speak of narrativity in music?

"Mamma sat down by my bed; she had chosen François le Champi, whose reddish cover and incomprehensible title gave it, for me, a distinct personality and a mysterious attraction. I had not then read any real novels. I had heard it said that George Sand was a typical novelist. This predisposed me to imagine that François le Champi contained something inexpressibly delicious. The narrative devices designed to arouse curiosity or melt to pity, certain modes of expression which disturb or sadden the reader, and which, with a little experience, he may recognize as common to a great many novels, seemed to me - for whom a new book was not one of a number of similar objects but, as it were, a unique person, absolutely self-contained - simply an intoxicating distillation of the peculiar essence of François le Champi. Beneath the everyday incidents, the ordinary objects and common words, I sensed a strange and individual tone of voice." (Marcel Proust, Remembrance of Things Past, 1st Part: Swann's Way).*

I choose to open this address with this remarkable text by Proust since it precisely expresses the problem which I would like to examine in the musical field: is there really, in novels and stories, a set of ways to express or conduct the flow? Ways which may allow anticipation of events which constitute the content of narration, and that the non-informed reader senses as creating expectations, fears, sadness or happiness, even if he is unable to recognize or identify them? Is there a structure that could be called proto-narrative, time-organizing in the successive alternation between tension and release, repetitions and variations, full, dense time and empty or dead time? "I sensed a strange and individual tone of voice...". A structure that exists before any narration, any story, an intuitive structure, which is sensed in the dynamics of the temporal flow, a structure whose essentially cognitive nature we have tried to demonstrate for a long time, but which also appears as a proto-organization of emotions and feelings.

I would like to examine this suggestion from a psychological point of view, and I started with Proust's text, to which I shall return all through this communication, since it presents the problem-with all the novelist's sensitivity and imagination-in a way that is exactly opposed and symetric to the way in which J.J. Nattiez considers the problem in a very famous article, entitled "Peut-on parler de narrativité en musique?", a title to which my own lecture is a friendly wink.

J.J. Nattiez starts with a very simple definition of a story or of narration, which he borrows from Chatman (1978): a story is a connecting, a putting-in-relation of "existents" and "events". Characters (existents) do acts or are victims of others' acts (events). A double chain connects them to each other: a linear chain which is that of time, or rather of the order of appearance of the characters and events, and a causal chain which links certain events to others, or the momentary or longer state of certain characters to certain events. J.J. Nattiez makes 2 fundamental remarks: the first is that a story supposes an explicitly designated content in the verbal language which refers to a fact or

* C.K. Scott Moncrieff & Terence Kilmarting (translation), Random House, N.Y., 1989
an imaginary reality. There must be a special kind of listening to the musical piece in order to construct a story from it, and this content exists in music only if the composer decides to explicitly transmit a message or if I, the listener, decide to search for a message therein. If the musical story were manifest, it would be the equivalent to the linear story, which is not the case. The second remark has to do with the fact that the story, even the literary one, simultaneously implies the author's intention to tell something, and the reader's desire to hear a story. A story which he may re-construct in his own way, since the literary story—no more than a hypothetical musical story—is a report on reality, and the written text by the author necessarily calls for an interpretation on the part of the reader.

Furthermore, if I return to Proust's text, I can see how the young Marcel understands but a little of the "story" by George Sand, and to what extent, at the same time, he re-constructs another mysterious course, with obscure zones, without the idea of a story or a "true novel" ever leaving his mind for an instant.

"The plot began to unfold: to me it seemed all the more obscure because in those days, when I read, I used often to daydream about something quite different for page after page. And the gaps which this habit left in my knowledge of the story were widened by the fact that when it was Mamma who was reading to me aloud she left all the love-scenes out. And so all the odd changes which take place in the relations between the miller's wife and the boy, changes which only the gradual dawning of love can explain, seemed to me steeped in a mystery the key to which (I readily believed) lay in that strange and mellifluous name of Champi, which invested the boy who bore it, I had no idea why, with its own vivid, ruddy, charming color."

This is, then, a story with "gaps", a mysterious story, the coherence of which does not appear to the young reader, simply because he finds more charm in the reverie or in his mother's voice and intonation than in the plot, which becomes but general movement, a minimal coherence, a sort of frame-form which reminds us that we are in a story and not in reality. Here, for the young Marcel, narration is essentially a continuity of voice, of rhythm, of intonative melody, a directing line vaguely perceptible through the particular quality of the name François le Champi, a representation which is sensitive to an internal emotional reality, in which the events are merely an excuse for the manifestation of this state in the reader's conscience. In short, narration is first of all an action of staging, a "conducting of a story", a dramatisation which removes us from reality (Freud mentioned that dreams are a dramatisation of ideas), from its "existents" and its "events", a strategy that both the author and the reader use for giving meaning. This may be what we should call "narrative impulse", to use J.J. Nattiez's expression. I would now like to address the question of what this act consists of psychologically.

In past works (Imberty, 1995 and 1997), I have shown the importance given by psychology today to temporal structuration phenomena in the development of the cognitive and affective life, starting from the beginning of life. We know, in particular, that the beginning of the infant's socialisation (between 3 and 6 months) is based on a repetitive organization of behaviors created by the mother in her relations with it: vocalisations, movements, tactile and kinesthetic stimulations are repeated without there being, at least at first, any pedagogical intention. In fact, in early age, repetition appears to be the mother's privileged mode to enter into a relation with her child since, the child's own repertory being limited, the maternal repetition fills a void where what is said or done has less importance than the stimulations' sensory quality, as well as their structuration. Concerning the mother and her child's vocal games, D. Stern* writes: "What is undoubtedly of less importance is what the mother actually says. What is

* personal translation
important, is the musicality of the sounds she produces. From this point of view, the repetitive action acquires its importance as a structural and functional unit in interaction" (1977, p.122). And the temporal structuration of interactions- or of interactive behaviors- which the repetitions engender, is no doubt the essential phenomenon on which our understanding of personality development must be based.

In her games with the infant, the mother is brought to repeat many things (actions, words,...). But it is clear upon observation that she never repeats them in the exact same way: "there is, says D. Stern, a progressive introduction of variations... The general form (of the repetitive action) can thus be conceptualized as a presentation and a representation of a theme with or without variations. More than half the repetitions, whether they are vocal or non-verbal, have variations in them" (ibid.). Thus, two elements allow the development of socialisation, affect and cognition in such situations: on the one hand, the infant learns to adapt to an ever-growing number of variations, but on the other hand, he can only do this because repetition is based on a regular rhythm which permits predictions and organizes time. The emotional alternation of tension and release, of insatisfaction and satisfaction, with their diverse transpositions, their different meanderings, is based on this regularity. In short, we see that the whole development of social and communicative behavior- and thus of personality- is constructed upon a learning of sequences whose temporal structure is based on repetition which permits the child to master time through varied regularity, ornamented and diversified regularity. We also find here what constitutes the universal substrate of music in all cultures, as well as story techniques.

There is probably a deep connection between affective experience and repetition: inasmuch as it structures time, repetition also structures the subject's emotional experiences. An example often cited by D. Stern is the game "got you". The mother varies the rhythm of her movements and words on the basis of a repetitive sequence which is almost isochronous, her voice growing in intensity. The child expects this isochronie, adopts, or interiorly experiences a regular tempo, but has to adjust his responses to the real succession of his mother's movements and words, which are put to rhythm in a different way, constituting a free variation on the basic regular pulse. It is thus irregularity due to variations in the mother's behavior which provokes the infant's participation through his own expectancies, but these only function because he himself is regulated on a latent isochronie which produces a playful disynchronisation between the true rhythm and the pulse. This principle of anticipating an event (in this case a movement, a word or a behavioral sequence) which is predictable but in a slightly indetermined way, may also be found at the basis of many games of tension in music.

Two new ideas appear: the first is that it is the mother who structures time by repetition-variation. From this structuration, the child slowly constructs and remembers an abstract model engraved in memory in an a-modal, global, syncretic way. What remains in the end is only a temporal contour, a course which is restricted to the subject's experience by a beginning and an end. For every interactive situation, we can easily imagine that the same process takes place, and very quickly the child acquires a very rich inventory of abstract temporal contours of behavior sequences à deux which are those which D. Stern calls "schemas of being-with", being with a person in such and such a situation, in such a context, in such an action, in such a game. The expression, that Stern uses since his last and important work The Motherhood Constellation (1995), emphasizes that all this is constructed only if there is human interaction. The second central idea which stems from this, and of which we may say that it allows a good understanding of musical behavior, is that of affective tuning. The notion is very complex, and I will only clarify it progressively. I will only mention for the moment that the notion of tuning is richer than that of adjusting: tuning, to tune in with the Other, that is to be on the same diapason, the same rhythm. The schema of being with would not function without tuning. And, as I have suggested, the notion includes the idea of a chord, in the musical, instrumental sense of the term: to tune in to someone in order to play together, not only to imitate, to tune in order to be together in tempo, in time, like the instrumentalists of a
The interesting hypothesis is then that the unity of an interpersonal or interactive experience is its temporal structure. It is upon this structure that the sensory, motor and affective experiences are added to constitute interiorised representations. In the very young child, these units of representation can not be more than a few seconds long, often much less, but they are subjected to the repetition of social interaction. Later on, they are progressively integrated into more complete representations, the organization of which stays at first connected to duration and to changes in duration, of levels and directions of psychic tension and release.

What is the profound nature of these phenomena? In his book from 1985, Le Monde Interpersonnel du Nourisson (1985, French translation 1989), D.Stern develops several interesting concepts which, as we shall see, are related to music.

The first of these is that of vitality affect, and we must give a rather long quotation of the text: "... many emotional characters do not fit in the existing lexicon or the taxinomy of affects. The unseizable characters are better rendered by dynamic, kinetic, terms such as "to appear", "to swoon", "fleeting", "explosive", "crescendo", "decrescendo", "burst", "to lengthen", etc. These characters are surely perceptible to the infant, and they are of a daily importance, even if it is only momentary". These vitality affects are then characters linked to emotions, to ways of being, to different ways of interiorly feeling emotions. They would be, for example, what distinguishes an "explosive" joy from a "fleeting" one, or the thousand ways of smiling, of getting up from one's chair, of taking a baby in one's arms, feelings which are not reducible to classic categorical affects, but which tone them in a way that is always sensitive to the individual.

If I were to translate D.Stern's idea in a different way, I would say that these feelings, are first of all of a temporal and dynamic nature, and that is what constitutes their originality. They give the instant, the action's or the emotion's present a thickness, and this is no doubt what the infant perceives first from acts, movements, attitudes expressed by his mother or people around him. These are ways of feeling, of being with, before being particular emotions or sentiments. The comparison with music or dance seems obvious, because the choreographer or the composer rather translate a way of feeling than a particular sentiment: when this is the case, what makes the stylistic originality, is precisely the fact that the abstract sentiment which is translated is never translated in the same way, that is never felt in the same way by two different composers.

I have myself touched upon this problem in my comparative works on Brahms and Debussy: the sadness which is provoked by the former's music is not that which is provoked by the latter's, they do not belong to the same universe of affective and semantic connotations (or of semantized affective connotations). There is something which clarifies the relations of musical expression and style: on one hand, the sadness evoked by Debussy's Des pas sur la neige is not the sadness of the Intermezzo op.119 n.1, the first possessing, as the experiments showed, something more static, fixed, immobile, something mortal, the second, on the other hand, possessing a certain vital fluidity which distances it from despair and brings it to nostalgia; but on the other hand, Debussy's sadness in this piece also possesses something which likens it to the light, ephemeral and ironic joy of La Danse de Puck or the Minstrels: I find the same global "quality" of organization of the auditory material, the same "quality" of contrasts, ruptures, rhythms, melodic figures. Sadness and irony, despairing solitude and light joy refer to an identical mode of "feeling", a same mode of forming the auditory flux, a same "intentionality", or at least, a same "schema of being with" the auditory material and with time; a same "schema of being in the world", a style. Isn't style but an architecture of
vitality affects? (Imberty, 1981, p. 13-14). The answer given by D. Stern is identical to my own: "During spontaneous behaviors, the domain of vitality affects is the equivalent of style in art" (1985). The affects of vitality modulate, "style", fixed and rigid behavior programs like walking, smiling, etc... All humans walk, but even when I'm unable to distinguish your features, from afar, I recognize you by the way you walk...

The notion corresponding to that of vitality affect in music is undoubtedly that which I have proposed, based on the experiments on the semantization of the musical experience, in an attempt to characterize the dynamic and temporal aspects of forms: the notion of dynamic vector. Dynamic vectors are musical elements which transport temporal significations or orientation, progression, diminution or intensification, of repetitions or returns. For example, in one of my experiments on La Puerta del Vino, by Debussy, I asked subjects, both musicians and non-musicians, to verbally describe what they heard, in real time, while listening. By what may be called the coda of the piece (bar 78), there is a sudden move to the high octave which is played pp. The answers indicate "serenity", "immobilization", "calm", "extinction", "effacement". The transposition to the high tones is then felt through this slowing down of time (although there is no objective slowing down in tempo in the chosen version), by this "immobilization" or this "serenity" which lightens the density of duration and of writing. The transposition to the high octave is then not only a change in register, but takes on the meaning of a kind of opening and immobilization of musical time, that the last arpeggio of the piece suddenly breaks. This perceived and felt change is thus a dynamic vector which orients the listener's perception, his expectancies, his internal representations. The quality of this orientation depends on what the dynamic vector refers to, the vector which is here assimilated to a set of the listener's vitality affects, of which he has, in audition, the immediate experience or revival.

The same kind of affirmation can be made for the piece as a whole, as well as for clarification of the different nature of global temporal progressions in La Puerta del Vino and of Intermezzo in b flat minor, op.118 n.6 by Brahms, chosen for this particular comparison. The answers given by the subjects clearly show the essentially different qualities of the linking of sequences which composes each one of the pieces. In Brahms, through a complex thematic progression, the answers concern processes in becoming, passages from one state to another - without these states being in themselves definable -, movements which last ("something is going to happen, birth, darkening, agony, etc."). and the transitions are marked by this evolutive slowness. In Debussy on the other hand, the contrasts and juxtapositions give rise to responses in which the movements evoked are brief, precise, abrupt, where the changes in states are instantaneous and surprising ("brutal, contortion, delirium, tension, cry, freer, more fluid, calm, immobilisation, etc."). This "immediate", "intuitive" character of the dynamic vectors' comprehension is the reason why, as I have already shown (1981, p.128), they do not have an objectivable semiological reality. They only constitute, in the subjects' perception, the references to an intention of meaning which disconnect from the form "in becoming" and from the style which overflows it. I shall later return to this intention of meaning.

Two points should be clarified: first, it must be noted that the vitality affects do not depend - for their translation into behavior - on any particular sensory mode. The type of "perception" that they represent is a-modal, which means that the child - later on the adult - spontaneously "translates" the "feeling" into one mode or another, indifferently, and even stays for the most part in a more confused or indetermined, more immediate state of perception. Many experimental results have shown that babies are perfectly capable of making the inter-modal transfers at a very early stage, and that the basis of their "knowledge" is not affected by one determined modality: thus, "certain properties of people and things, such as form, level of intensity, movement, number and rhythm, are directly apprehended as global and a-modal perceptive attributes" (Stern, 1985).
Later, the vitality affects, being "un-categorizable" since they are a-modal, find their consistence in their "activation profile", their internal temporal schema. Francès, in 1958, showed that musical expressiveness is based on the provocation of "sentimental abstracts" which would be "the echos and the shadows of the sedimented corporal experience" (1958). The hypothesis is interesting, but on the one hand, musical expressiveness, as we can see, cannot be reduced to a provocation of just the corporal experience, even if this experience does not seem negligible, and on the other hand, the term "abstract" seems not very well-chosen, since it seems to imply that cognitive abstraction is indispensable for the functioning of a "recognizing" or an "identification" of the formal schemas in music. Conversely, the concept of vitality affect is rooted in the dynamics of affective life, in the dynamics of personality emerging and of interpersonal relations, in what Bergson would call the creativity of the vital élan. The vitality affect is thus itself an emerging time, a fragment of time in the present which is sensed as a sequence of tensions and releases of varying strength, like a sequence of variations in sensation intensity. An enlightening and remarkable example given by D. Stern: in trying to calm her baby, the mother will tell him "Hush now, hush now...". She does this by emphasizing the first syllable, slowing down on the second. But she can try to obtain the same effect without saying anything, by just caressing the child's head: the movement, the caress, have thus the same profile, accentuated at the beginning, slower and lighter at the end. What is interesting here is that the baby feels the two behaviors in the same way, and thus experiences the same vitality affect (1985) which is characterized by a determined activation profile which he feels and "recognizes" immediately. As experiences of the same type accumulate, the vitality affects are grouped into very global organizations, in which the perceptions, acts and thoughts do not exist as such, but these organizations constitute the "matrix" of their development in later experiences. Thus, vitality affects may be considered as the primitive categories (sensitive and intuitive) on which later on during development, emotions, sentiments, perceived and identified forms, thoughts, will be constructed. Their organization, says D. Stern, remains "the fundamental domain of human subjectivity", it is "the fundamental reservoir in which we may draw all the experiences of creation" (ibid).

Two new notions appear, the first of which brings us back to the experience of time and to the creation of musical time, the other to the very controversial question of semiotisation of musical time and of the forms which it takes in different musical pieces and in styles, in a sense narrativity in music.

First, the notion of temporal feeling shape (or temporal web of feeling) seems to derive naturally from all that precedes: it is defined in The Motherhood Constellation (1995) as a form of representation of the affective experience. It is, then, a contour of affectivity, the temporal form of a set of intensity, rhythm and duration profiles of vitality affects, the coherence of which it assures for the subject, in a durable present. Its emergence is, for the subject, an event produced in real time, "within" experience. But in order to understand the range, the bearing of this notion, we must ask: what is being shaped? In other words, what is woven, as a plot is, an event about which we expect a story, that is that we wait to be given the meaning of? Again put in a different way, if it is "woven", what meaning does it have?

Behind this question lies a much more general question about time: it only has meaning for us because it is a link, a passage between past, present and future, it has an orientation in which our life is projected. In babies, it is clear that existential time is still fragmented, that general orientation does not exist, that it is reduced to the present, that is to sensations, to movements, to interpersonal actions and related vitality affects. In the beginning - we saw this previously - it may just be a few seconds long, or even less. But in this brief present, how is the coherence of self and the world constructed, how, without the intuition of a future temporal horizon, can the subject pick up the direction of what is weaving itself in the felt temporal flux of time? Many philosophers have evoked what they call the viscosity of the present, the extension of present into the time of
subjective feeling. Saint-Augustin distinguished between the units of objective time and those of subjective time, which expand as a function of internal experience, in such a way that it is then possible to speak of a past of the present, a present of the present and a future of the present: "Time is a distension of the soul", the future is only the anticipation of the future, the past only the memory of the past, the present the anticipation, itself concentrated on the object that is not yet and which in an instant shall be no more.

What we find in Saint-Augustin, is this issue of meaning, not only of the directionality of time and of what the subject experiences of it, but "where is it going?" as an intentionality of sensations, of perceptions, of acts, of representations. The present is going towards something because consciousness of time is consciousness of something in time, consciousness of some goal to attain, as observed by Husserl as well. Time unifies, duration constitutes the link between feelings beyond their heterogeneity, it is co-substantial to the emerging of the self. What is this link? Or, what window of time permits this coherence, this meaning? This is the same as asking how we can have temporal representations. In the baby, the capacity to find coherence in the interpersonal world is linked to its capacity to feel the link between its actions and the pleasure or displeasure that it feels. In other words, the actions are colored by an intrinsic motivation - not stated as a conscious goal, the pleasure or displeasure being themselves immediate feelings - which maintains the energy to act. All behaviors of auto-excitation stem from this mechanism. Upon it, the mother's behavior is constructed and tuned. The "experience of doing" is colored by an orientation, by a goal, which, a posteriori gives this experience a coherence through a precise limitation in time: to do in order to feel pleasure, then little by little in response to the other, to share with the other; but once the goal is reached, the action is centered again on another sequence, the preceding one being sent into the past, momentarily forgotten. It is this sending into the past which, a posteriori, gives the sequence coherence, makes it appear as a temporal form having had a beginning, a middle and an end.

This is what D.Stern calls "the proto-narrative envelope" (1994 and 1995): in fact, the narrative form, in a universe of language and signs to which the baby will have access later on, is what constructs the unity of time, shapes the reality of the human development (becoming). The temporal "after-the-fact" is thus a semiotisation of the vitality affects' activation profiles, or more precisely, it is what allows semiotisation to develop in duration, which gives form to the temporal feeling shapes, in sum, it is what makes something "weave itself", "take meaning" in time. The proto-narrative envelope is thus an affectivity contour spread out in time with the coherence of a quasi-intrigue: it is organized around the putting-into-act of an intention-motivation (orientation towards a goal), it defines a portion of time in which the baby feels its own coherence, that is simultaneously brings to himself (to his sense of a core self) the sensations and the needs (for example, hunger), of his acts (movements, cries...), of his perceptions (face, caresses, mother's voice...), of feelings (vitality affects linked both to his sensations, his movements, his perceptions), but it stays, beneath any language, a line of intuitive dramatic tension.

The proto-narrative envelope is then a proto-semiotic form of interior experience of time, a matrix of the "story" of tensions and releases linked to the "plot", "intrigue" (or the "proto-intrigue", 1994) of the search for satisfaction, it is what gives experience its global unity, whatever its degree of complexity. This unity is "experienced" through the baby's progressively acquired consciousness of causality and of the advancing of experience towards its final state. Around 3 or 4 months of age, the baby becomes capable of differentiating itself from the other, the self being sensed as the author or the cause of deliberate actions. There is then an intuitive perception of an agent-self, of an action, of an instrument and of a goal. These elements are at the basis of primitive forms of causality, they organize proto-intrigues, and bring out of the temporal feeling shapes a line of dramatic tension which orients it: "When the motivation (desire) is activated in an interpersonal situation, it subjectively creates a quasi-narrative structure" (1994).
We arrive here to what P. Ricœur says in *Temps et récit* (1983) when he discusses the "pre-narrative structure of experience". Our experiences, in the dynamics of what Bergson would call the vital étan, are felt as story courses, webs, which little would be needed for them to be told and whose coherence would knot itself in the act of narration. Indeed, little would be needed? A coding in language, a distance to reality created by its representation, a temporality that is mastered and, most of all, someone to tell it to, be that someone ourselves.

These facts and reflections bring a new light to the concept of the musical piece's *macro-structure* which I have developed on several occasions (1981, pp. 88-90; 1991). I have defined macro-structure as a schema of the structuration of time, that is as the schema of progression of temporal structures of tension and release of the musical piece. In other words, a musical piece is first of all an ordering of auditory events in time, the macro-structure a simplified schema type, an a-priori ordering that later the concrete auditory events come to fill in, events whose progression for the listener may be defined as a structured and hierarchized sequence of tensions and releases. The macro-structure is defined simultaneously at the levels of musical grammar, of cognitive operations in composition and understanding of a piece, but also on the level of the expressivity of the listener's feelings during audition. The temporal progression, through tensions and releases, through the formal patterns which evoke what I can now name vitality affects, takes meaning in this oriented continuity from the beginning to the end of the piece, finds its coherence in this temporal course which links melodic, rhythmic, harmonic movements, strings the "proto-narrations" of a thousand nuances of the interior life into a unique profile. In commenting the text of *Essai sur les données immédiates de la conscience* by Bergson, V. Jankélévitch describes this well: "But what middle way can ever link pain and joy? And yet duration does this miracle". Interior duration, as harmonic modulation, implies "the intuition of a certain thickness of originalities to surpass." (1959, pp. 4-43).

Thus, the macro-structure clearly appears as the organized *temporal course of feelings or experiences*, vitality affects designed in their proto-narrative profiles, transported in the flux of musical time. Music is thus like the translation or the representation of the original matrix of all symbolic forms, of all forms of language, that is of all forms of ordering of time. From a developmental psychological and a psychoanalytic point of view, we can understand the profound intuition of Lévi-Strauss in *L'Homme Nu*: "in coding myth into sounds instead of into words, the musical piece supplies a key to decoding, a matrix of relations which filters and organizes lived experience, substitutes itself to it and procures the salutary illusion that contradictions can be overcome and difficulties resolved" (1971, p.590).

Would music then also be the original matrix of all story forms? J.J.Nattiez is right in forming the essential question: isn't the idea of narration applied to the musical piece an indication of the confusion that the contemporary listener feels when faced with pieces which no longer possess anything linear? Thus pieces in which "it does not weave", where nothing happens, or where nothing progresses? I concur with the two hypotheses which are most relevant in C.Deliège (1989): starting from Debussy, the internal organization of sequences which compose a musical piece is more important than the discursive process which links them in the whole form (and this corresponds perfectly with what Debussy was looking for in order to break any symmetry, to establish in music the primacy of the instant in relation to the lasting process, a search that contemporary composers have prolonged until arriving at the sound-sequence); however, the listener's demand to find a minimum of clarity and intelligible order in the musical piece, appears contradictory with the weakening, the disappearance of the discursive. In cognitive psychology terms, this demand is in fact the possibility to construct perceptive *anticipation* and *retroactions* without which no unity can be perceived or
conceptualized, neither in music nor in any other domain of human activity or knowledge.

A whole portion of contemporary music, especially serial music, gives preference to the small unitary sequences of time, the elementary proto-narratives, these dynamic vectors of which we spoke earlier, to the detriment of the progression of the whole, of the temporal movement encompassing the piece in its integreality. Serial music is constructed on these minimal temporal adventures which remain without connection. For at the same time, the principle of non-repetition on which Schoenberg founds his system, makes it improbable that the temporal course englobing these minimal sequences defined around the series may ever take meaning for the listener. The "grand form" stays purely abstract, it does not link the proto-narrative envelopes between them in a re-presentation of lived experience, it stays foreign to the course of interior duration and of its possible symbolization in the time of the complete piece.

Since basically this music shuns repetition, it engenders this smooth time of which Boulez (1963) spoke, the time of the instant, time which does not pass, time which can no longer be the time of an adventure and that the listener feels sometimes as the time of boredom. The refusal of any expressivity, the refusal of any semantization of music in the years of serialism is thus not pure chance in a psychological sense: it is connected to the destruction of the "grand form", that is to the destruction of any thought of time, of any thought of a defined trajectory that is predictable in time: the ultimate result of this tendancy of musical art in the first half of the XXth century is undoubtedly the Momentform, that is a form which has nothing linear and which is entirely centered around the material. The absence of any discursive process makes it then totally unpredictable and indiscernible. And, if from the three Vienna men, Webern seems to us the most enigmatic as weil as the most subtly innovative, it is because, with him, even though variation always gets the better of repetition, he has nonetheless achieved an exceptional equilibrium between the minute-discursive and the sequence-piece!

"And so, when she read aloud the prose of George Sand..taking pains to banish from her voice any pettiness of affectation which might have choked that powerful stream of languatd, she supplied all the natural tenderness, all the lavish sweetness which they demanded to sentences which seemed to have been composed for her voice and which were all, so to speak, within the compass of her sensibility. She found, to tackle them in the required tone, the warmth of feeling which pre-existed and dictated them, but which is not to be found in the words themselves, and by this means she smoothed away, as she read, any harshness or discordance in the tenses of verbs, endowing the imperfect and the preterite with all the sweetness to be found in generosity, all the melancholy to be found in love, guiding the sentence that was about to begin, now hastning, now slackening the pace of the syllables so as to bring them, despite their differences of quantity, into a uniform rhythm, and breathing into this quite ordinary prose a kind of emotional life and continuity".

I end with this text by Proust: it supplies me, I believe, with an enlightening conclusion: we know that it is through this episode and its strong affective context (going to bed which separates the child from his mother, the crying fit and the reading of François le Champi", "a real novel") that Proust enters into literature. The text's movement is fundamentally the inverse of the procedure adopted by those who would at any cost apply the categories and theories of narration to music. Gradually, what becomes essential, is all that is not the story, so mysterious that it doesn't arouse the attention of the young Marcel to its actual content, it is the elan of the phrase, the rhythm of the voice, its particular music that constitute the unconscious essence of the proustien phrase, it is the temporal course of reading, not the logic of the events. Mrs. Proust recreates for her son, and without her knowing it, the proto-narrative envelope of the novel, that is what is underneath the novel, what is BEFORE the novel, the particular
emotion which conditions only this particular story (but it could have been another) will have this impact on the formation of the young adolescent's personality. But it is also what allows him, in this particular experience, to construct his internal coherence, to link his past in Combray to his vocation as an author.

In the same way, in the analysis of semiological structures of music, we first have to find this movement, this dynamics of temporal linearity, this "intrigue" without characters or events, these fractures and these restorations of the coherence of our becoming. We cannot talk about narrativity in music, since music does not tell, does not represent. No more than the voice of Mrs. Proust tells to her son. She only stages time and our ways of feeling it, of reconstructing it, of fantasizing it between continuity and discontinuity, between fusional unity and division, between mobility and immobility. The story of musical creation is the story of the relations of man to time, both on the individual and on the collective level. But the course through the remarkable works of D. Stern suggests a new depth: the constant reference to the fact that individual psychology of time is forged in the interactions with others, the major assertion that the sentiment of duration is born in the game of interactive communication, make us understand that music only touches us through the other. I am totally of the opinion of J.J. Nattiez when he shows that "musical communication" (in the banal sense) is but an illusion. But I think that the problem of musical expression lies elsewhere: music draws its power in its profoundly social nature, just like language as a vehicle of interiorised representation. All the temporal substance of music is fed by our ways of being in the world, that is of being in our time, our culture, with our perception, our body, our emotions and our sentiments. It is not communication, but it is a representation of our capacity to communicate, it is a stylised game of our opening to the world, it is communication without an object to communicate. It is proto-narration, that is, literaly before narration, irredicuble to its categories and its logic decompositions of reality.

BIBLIOGRAPHIE


Delliège, C. (1989), La forme comme expérience vécue, In S. McAdams & I. Deliège (eds.), La Musique et les Sciences Cognitives (pp. 159-179), Bruxelles, Mardaga.


Stern, D. N. (1977), Mère et enfant. Les premières relations, Bruxelles, Mardaga.


Peut-on parler sérieusement de narrativité en musique ?

"Maman s'assit à côté de mon lit; elle avait pris François le Champi à qui sa couverture rougeâtre et son titre incompréhensible donnaient pour moi une personnalité distincte et un attrait mystérieux. Je n'avais jamais lu encore de vrais romans. J'avais entendu dire que George Sand était le type du romancier. Cela me disposait déjà à imaginer dans François le Champi quelque chose d'indéfinissable et de délicieux. Les procédés de narration destinés à exciter la curiosité ou l'attendrissement, certaines façons de dire qui éveillent l'inquiétude et la mélancolie, et qu'un lecteur un peu instruit reconnaît pour communs à beaucoup de romans, me paraissaient simplement - à moi qui considérais un livre nouveau non comme une chose ayant beaucoup de semblables, mais comme une personne unique, n'ayant de raison d'exister qu'en soi - une émanation troublante de l'essence particulière à François le Champi. Sous ces événements si journaliers, ces choses si communes, ces mots courants, je sentais comme une intonation, une accentuation étrange. ..." (Marcel Proust, À la recherche du temps perdu, 1re partie : Du côté de chez Swann ).

Si je commence ma conférence par cet admirable texte de Proust, c'est qu'il pose très exactement le problème que je voudrais Examiner pour la musique : y a-t-il réellement dans les romans et les récits un ensemble de façons de dire ou de conduire la trame qui se devine, voire permettre d'anticiper sur le temps des événements eux-mêmes qui font le contenu de la narration, et que le lecteur non averti, même s'il ne les reconnaît pas et ne les identifie pas, ressent comme créant les attentes, les angoisses, la tristesse ou la joie ? Y a-t-il en somme une structure que l'on pourrait dire proto-narrative, organisatrice de temps en alternances successives de tensions et de détentes, de répétitions et de variations, de temps pleins et denses et de temps vides ou morts ? "Je sentais comme une intonation, une accentuation étrange..." Structure d'avant toute narration, tout récit, structure intuitive et intuitivement ressentie dans la dynamique de l'écoulement temporel, structure dont on a longtemps cherché à démontrer la nature essentiellement cognitive, mais qui se manifeste aussi comme une proto-organisation des émotions et des ressentis.

Voilà la proposition que je voudrais examiner du point de vue de la psychologie, et si je suis parti du texte de Proust auquel je reviendrai d'ailleurs tout au long de cet exposé, c'est en un sens parce qu'il présente le problème - avec toute la sensibilité et l'imagination de l'écrivain - de façon exactement opposée et symétrique à la façon dont J.J. Nattiez le considère dans un article fort célèbre, intitulé "Peut-on parler de narrativité en musique ?", titre auquel celui de ma propre conférence est un amical clin-d'œil.

J.J. Nattiez part en fait d'une définition très simple du récit ou de la narration qu'il emprunte à Chatman (1978) : un récit est une mise en relation d'"existants" et d'"événements". Des personnages (existants) accomplissent des actes ou sont victimes d'actes d'autres
(événements). Une double chaîne les relie les uns aux autres : une chaîne linéaire qui est celle du temps, ou plutôt de l'ordre d'apparition des personnages et des événements, une chaîne causale qui relie certains événements à d'autres ou l'état momentané ou plus ou moins durable de certains personnages à certains événements. De tout ceci, J.J. Nattiez tire deux remarques fondamentales : la première est qu'un récit suppose bien un contenu désigné explicitement dans le langage verbal qui renvoie à une réalité de fait ou imaginaire. Il faut une écoute particulière de l'œuvre musicale pour construire un récit à partir d'elle, et ce contenu n'existe en musique que si le compositeur a décidé de transmettre explicitement un message ou si j'ai décidé, moi auditeur, d'y rechercher un message. Si le récit musical était manifeste, il serait alors équivalent au récit littéraire, ce qui n'est pas le cas. La seconde remarque tient au fait que le récit, même littéraire, implique bien à la fois l'intention de l'auteur de raconter quelque chose, et le désir du lecteur d'y entendre une expansion. Histoire qu'il lui arrive alors de re-construire à sa guise, parce que le récit littéraire - pas plus que ne le serait un hypothétique récit musical - n'est un reportage sur la réalité, et que le texte écrit par l'auteur appelle nécessairement une interprétation de la part de celui qui le lit.

D'ailleurs, si je reviens au texte de Proust, je m'aperçois combien le jeune Marcel ne comprend que peu le "récit" de George Sand, et combien, en même temps, il re-construit une autre trame mystérieuse, avec ses zones d'ombre, sans que pour autant l'idée-même de récit - ou de "vrai roman" - ne quitte un instant son esprit.

"L'action s'engagea; elle me parut d'autant plus obscure que dans ce temps-là, quand je lisais, je rêvais souvent pendant des pages entières à tout autre chose. Et aux lacunes que cette distraction laissaient dans le récit, s'ajoutait, quand c'était maman qui me lisait à haute voix, qu'elle passait toutes les scènes d'amour. Ainsi tous les changements bizarres qui se produisent dans l'attitude respective de la ménagère et de l'enfant et qui ne trouvent leur explication que dans les progrès d'un amour naissant me paraissaient empreints d'un profond mystère dont je me figurais volontiers que la source devait être dans le nom inconnu et si doux de "Champi" qui mettait sur l'enfant qui le portait sans que je susse pourquoi, sa couleur vive, empourpée, et charmante."

Voici donc un récit "à trous", un récit mystérieux dont la cohérence n'apparaît guère au jeune lecteur, simplement parce que celui-ci trouve finalement plus de charme à la rêverie ou à la voix et l'intonation de sa mère qu'à l'intrigue qui ne devient en somme qu'un mouvement général, une cohérence minimale, une sorte de forme-cadre qui vient rappeler qu'on est bien dans le récit et non dans la réalité. Ici, pour le jeune Marcel, la narration est essentiellement une continuité de voix, de rythme, de mélodie intonative, un fil directeur vaguement perceptible à travers la qualité particulière du nom de François le Champi, une représentation sensible d'une réalité interne émotionnelle, affective, dans laquelle les événements ne sont plus que prétextes à la manifestation de cet état à la conscience du lecteur. Bref, la narration est d'abord un acte de mise en scène, une "conduite de récit", une dramatisation qui éloigne de la réalité (Freud disait justement que le rêve est une dramatisation de l'idée), de ses "existants" et de ses "événements", une stratégie de l'auteur comme du lecteur pour donner sens. C'est peut-être ce qu'il faut appeler une "impulsion narrative", selon l'expression de J.J. Nattiez, et c'est la question de savoir si, psychologiquement, elle consiste que je voudrais examiner maintenant.

Dans des travaux antérieurs (Imberty, 1995 et 1997), j'ai montré l'importance qu'accorde la psychologie aujourd'hui aux phénomènes de structuration temporelle dans le développement de la vie cognitive et affective dès le début de la vie. On sait en particulier que tout le début de la socialisation de l'enfant (entre 3 et 6 mois) est basé sur une organisation répétitive des conduites créée par la mère dans ses relations avec lui : vocalisations, mouvements, stimulations tactiles et kinesthésiques sont répétées sans qu'au moins au début, il y ait intention pédagogique. En fait, en bas-âge, la répétition apparaît comme le mode privilégié de la mère pour entrer en relation avec son enfant, parce que, le répertoire propre de l'enfant
étant limité, la répétition maternelle vient combler un vide où ce qui est dit ou fait a moins d'importance que la qualité sensorielle des stimulations ainsi que leur structuration. A propos des jeux vocaux de la mère et de son enfant, D. Stern écrit : "Ce qui sans doute importe moins, c'est ce que la mère dit réellement. L'important, c'est la musicalité des sons qu'elle produit. De ce point de vue, l'action répétitive acquiert son importance en tant qu'unité structurelle et fonctionnelle dans l'interaction." (1977, p.122). Et la structuration temporelle des interactions - ou des comportements interactifs - qu'engendrent les répétitions est sans doute le phénomène essentiel sur lequel doit s'appuyer notre compréhension du développement de la personnalité.

Dans ses jeux avec l'enfant, la mère est donc amenée à répéter beaucoup de choses (actions, paroles...). Mais il est évident à l'observation qu'elle ne répète jamais exactement de la même façon : "il y a, dit D. Stern, introduction progressive de variations... La forme générale (de l'action répétitive) peut donc être conceptualisée en tant que présentation et re-présentation d'un thème avec ou sans variations. Plus de la moitié des répétitions, qu'elles soient vocales ou non verbales, comptent des variations."(ibid ). Deux éléments permettent dès lors le développement de la socialisation, de l'affect et de la cognition dans de telles situations : d'une part, l'enfant apprend à s'adapter à un nombre toujours plus important de variations, mais d'autre part, il ne le peut que parce que la répétition est basée sur un rythme régulier qui rend prévisible et organise le temps. C'est sur cette régularité que se fonde l'alternance émotionnelle de la tension et de la détente, de l'insatisfaction et de la satisfaction, en même temps que leurs diverses transpositions, leurs divers contournements. En bref, on voit que tout le développement du comportement social et communicatif - et donc de la personnalité - est construit sur l'apprentissage de séquences dont la structure temporelle est basée sur la répétition qui permet à l'enfant de maîtriser le temps par la régularité variée, la régularité ornée et diversifiée. Nous retrouvons bien là aussi ce qui constitue le substrat universel de la musique dans toutes les cultures, aussi bien que des techniques de récit.

Il y a donc probablement un lien profond entre les expériences affectives et la répétition : en tant qu'elle structure le temps, la répétition structure aussi les expériences émotionnelles du sujet. L'exemple que donne très fréquemment D. Stern est le jeu de "je t'attrappe". La mère varie le rythme de ses gestes et de ses paroles sur la base d'une séquence répétitive virtuellement isochrone, allant crescendo dans la voix. L'enfant attend cette isochronie, adopte donc ou vit intérieurement un tempo régulier, mais doit ajuster ses réponses à la succession réelle des gestes et des paroles de sa mère rythmées de manière différente, variation libre en somme autour de la pulsation régulière de base. C'est donc l'irrégularité due aux variations de la mère qui provoque la participation de l'enfant à travers ses propres attentes, mais celles-ci ne fonctionnent que parce qu'il est lui-même réglé sur une isochronie latente qui produit le décalage ludique entre le rythme réel et la pulsation. Ce principe de l'attente d'un événement (ici d'un geste, d'une parole ou d'une séquence comportementale) prévisible mais de manière légèrement indéterminée se retrouve aussi à la base de bien des jeux de tension musicale auxquels nous convient les compositeurs.

Deux idées nouvelles apparaissent ici : la première est que c'est la mère qui structure le temps par la répétition-variation. De cette structuration, l'enfant peu à peu construit et retient un modèle abstrait engrammé en mémoire de manière a-modal, globale, syncrétique. Il n'en reste en somme qu'un contour temporel, un parcours qui se délimite dans l'expérience du sujet par un début et une fin. Pour chaque situation interactive, on imagine facilement que le même processus se déroule, et très vite, l'enfant acquiert un stock très riche de contours temporels abstraits de séquences comportementales à deux qui sont ce que D. Stern appelle des "schémas-d'être-avec ", être avec une personne dans telle et telle situation, dans tel contexte, dans tel action, dans tel jeu. L'expression, que Stern utilise depuis son dernier et important ouvrage, The Motherhood Constellation (1995), marque bien que tout ceci ne se construit que s'il y a interaction humaine. D'où la seconde idée qui est centrale, et dont nous pouvons penser qu'elle permet de comprendre bien des comportements musicaux, celle d'accordage affectif. La notion est complexe, et je ne l'éclairerai que progressivement. Je retiendrai seulement que la notion d'accordage est plus riche que celle d'ajustage : accorder, s'accorder avec l'Autre, c'est-à-dire être sur le même diapason, le même rythme. Le schéma-d'être-avec ne fonctionnerait pas sans accordage. Et puis, je viens de le suggérer, la notion comporte bien l'idée d'accord, au
sens musical, instrumental du terme : s'accorder pour jouer ensemble, non pas seulement pour
s'imiter; s'accorder pour être ensemble dans le tempo, dans le temps, comme le font les
instrumentistes du quatuor qui se mouilent dans un schéma-de-jouer-avec sans lequel leur
entreprise serait vouée à l'échec.

L'hypothèse intéressante est donc ici que l'unité d'une expérience interpersonnelle ou
interactive est sa structure temporelle sur laquelle se greffent les expériences sensorielles,
motrices et affectives pour constituer des représentations intériorisées. Chez le très jeune enfant,
ces unités de représentations ne peuvent guère dépasser quelques secondes, souvent beaucoup
moins, mais elles sont soumises à la répétition de l'interaction sociale et, par la suite, s'intègrent
ainsi progressivement dans des représentations plus complètes dont l'organisation reste d'abord
liée à la durée et aux changements dans la durée, des niveaux et des directions des tensions et
des détentes psychiques.

Quelle est la nature profonde de ces phénomènes ? Dans son livre de 1985, Le monde
interpersonnel du nourrisson (1985, trad. frisée 1989), D. Stern développe plusieurs concepts
intérressants dont nous allons voir qu'ils ont quelque rapport avec la musique.

Le premier d'entre eux est celui d'affect de vitalité, et il faut citer un peu longuement le
texte : "...de nombreux caractères des émotions ne rentrent pas dans le lexique existant ou dans
la taxinomie des affects. Ces caractères insaisissables sont mieux rendus par des termes
dynamiques, kinétiques tels que "surgir", "s'évanouir", "fugace", "explosif", "crescendo",
"descrescendo", "éclater", "s'allonger", etc. Ces caractères sont certainement perceptibles par le
nourrisson, et d'une importance quotidienne, même si elle n'est que momentanée." (1985, trad.
frisée p.78). Ces affects de vitalité sont donc des caractères liés aux émotions, aux façons
d'etre, aux diverses façons de ressentir intérieurement les émotions. Ce sera par exemple tout ce
qui sépare une joie "explosive" d'une joie "fugace", ou bien ce sera encore les mille façons de
sourire, de se lever de sa chaise, de prendre le bébé dans ses bras, ressentis qui ne sont pas
réductibles aux affects catégoriels classiques, mais qui viennent les colorer de manière toujours
très sensible pour le sujet.

Si je traduis autrement l'idée de D. Stern, je dirai que ces ressentis sont d'abord de
nature dynamique et temporelle, que c'est là ce qui fait leur originalité. Ils donnent une
épaisseur à l'instant, au présent de l'action ou de l'émotion en cours, et c'est sans doute
d'abord cela que perçoit le nourrisson des actes, des gestes, des attitudes de sa mère ou des
personnes qui l'entourent. Ce sont des façons de sentir, d'être avec, avant d'être des émotions
ou des sentiments particuliers. La comparaison avec la musique ou la danse s'impose alors, car
le chorégraphe ou le compositeur traduisent bien plus une façon de sentir qu'un sentiment
particulier : lorsque c'est le cas, ce qui fait l'originalité stylistique, c'est précisément que le
sentiment abstrait traduit n'est jamais traduit de la même façon, c'est-à-dire ressenti de la même
façon par deux compositeurs différents.

J'ai moi-même longuement évoqué ce problème dans mes travaux comparatifs entre
Brahms et Debussy : la tristesse qu'évoque la musique de l'un n'est pas la tristesse qu'évoque
la musique de l'autre, elles n'appartiennent pas au même univers de connations affectives et
sémantiques (ou de connotations affectives semantisées). Il y a là quelque chose qui éclaire les
rapports de l'expressivité musicale et du style : d'une part, la tristesse qu'évoque Des pas sur la
neige de Debussy n'est pas la tristesse de l'Intermezzo op. 119, n°1, la première possédant,
comme le montrent les expériences, quelque chose de plus statique, figé, immobile, quelque
chose de mortel, la seconde possédant au contraire une certaine fluidité vitale qui l'éloigne du
désespoir et la porte vers la nostalgie; mais d'autre part, la tristesse de Debussy dans cette pièce
possède aussi quelque chose qui la rapproche de la joie légère, éphémère et ironique de La
Danse de Puck ou des Minstrels : j'y découvre la même "qualité" globale d'organisation du
matériau sonore, la même "qualité" des contrastes, des ruptures, des rythmes, des figures
mélodiques. Tristesse et ironie, solitude désespérée et joie aérienne et bondissante renvoient à
un même mode de "ressentir", un même mode de façonner le flux sonore, une même "intentionnalité", ou tout au moins, un même "schéma d'être avec" le matériau sonore et le temps, c'est-à-dire un même schéma d'"être au monde", un style. Le style ne serait-il donc qu'une architecture des affects de vitalité ? (Imberty, 1981, pp.13-14). La réponse de D. Stern est en tout cas identique à la mienne : "Au cours des comportements spontanes, le domaine des affects de vitalité est l'équivalent du style dans l'art." (1985, trad.frçèse, p.206). Les affects de vitalité modulent, "stylisent" des programmes comportementaux fixes et rigides comme la marche, le sourire, etc...Tous les humains marchent, mais, alors que je ne distingue pas encore vos traits, au loin, je vous reconnais à votre démarche...

La notion correspondant à celle d'affect de vitalité en musique est sans doute celle que j'ai proposée à partir des expériences sur la sémantisation de l'expérience musicale pour caractériser les aspects dynamiques et temporels des formes : il s'agit de la notion de vecteur dynamique. Les vecteurs dynamiques sont des éléments musicaux qui véhiculent des significations temporelles d'orientations, de progression, de diminution ou de croissance, de répétitions ou de retours. Par exemple, dans l'une de mes expériences sur La Puerta del Vino, toujours de Debussy, j'ai demandé à des sujets musiciens et non musiciens, de décrire verbalement ce qu'ils entendaient, en temps réel au cours de l'audition. Or, à ce qu'on peut appeler la coda de la pièce (mesure 78), il y a un brusque passage à l'octave aigüe qui se fait pp. Les réponses indiquent "sérénité", "immobilisation", "calme", "extinction", "effacement". Le passage à l'aigu est donc ressenti dans ce ralentissement du temps (bien qu'il n'y ait pas de ralentissement objectif du tempo dans l'interprétation choisie), par cette "immobilisation ou cette sérénité" qui détend la densité de la durée et de l'écriture. Le passage à l'octave aigüe n'est donc pas seulement un changement de registre, mais prend la signification d'une sorte d'ouverture et d'immobilisation du temps musical, que le dernier arpège de la pièce vient subitement briser. Ce changement perçu et ressenti est donc un vecteur dynamique qui oriente la perception de l'auditeur, son attente, ses représentations internes. La qualité de cette orientation dépend de ce à quoi renvoie le vecteur dynamique, assimilé ici à un ensemble d'affects de vitalité dont l'auditeur fait immédiatement, à l'audition, l'expérience ou la reviviscence.

Le même genre de constatations peut être fait pour l'ensemble de la pièce, de même qu'il peut être fait pour éclairer la nature différente des progressions temporelles globales de La Puerta del Vino et de l'Intermezzo en si bémol mineur, op.118 n°6, de Brahms choisi pour la comparaison. Les réponses données par les sujets font nettement apparaître les qualités très différentes des enchaînements des séquences qui composent chacune des œuvres. Chez Brahms, à travers une progression thématique complexe, les réponses concernent des devenirs, des passages d'un état à un autre - sans que ces états soient eux-mêmes précisés - des mouvements qui durent ("quelque chose va arriver, naissance, assombrissement, agonie, etc."), et les transitions sont marquées de cette lenteur évolutive. Chez Debussy au contraire, les contrastes et les juxtapositions donnent lieu à des réponses où les mouvements évoqués sont brefs, précis, brusques, où les changements d'état sont instantanés et surprenants ("dénanchement brutal, délire, tension, cri, plus libre, plus flou, calme, immobilisation, etc."). Ce caractère "immédiat", "intuitif" de la compréhension des vecteurs dynamiques est la raison pour laquelle, comme je l'ai déjà montré (1981, p.128), ils n'ont pas de réalité sémiologique objective. Ils constituent seulement, dans la perception des sujets, les repères d'une intention de sens qui se dégage de la forme en devenir et du style qui la déborde de toutes parts. Sur cette intention de sens, je reviendrai plus loin.

Deux points doivent être précisés : tout d'abord, il faut noter que les affects de vitalité ne dépendent - pour leur traduction dans les comportements - d'aucun mode sensoriel particulier. Le type de "perception" qu'ils représentent est a-modal, ce qui signifie que l'enfant - plus tard, l'adulte - "traduit" spontanément le "ressenti" dans un mode ou un autre, indifféremment, et s'en tient même le plus souvent à un état de "perception" plus confus ou indéterminé et plus immédiat. Bien des résultats expérimentaux ont montré que les bébés sont parfaitement capables d'effectuer les transferts inter-modaux très tôt, et que la base de leurs "connaissances" n'est pas affectée d'une modalité déterminée : ainsi, "certaines propriétés des personnes et des choses, telle que la forme, le niveau d'intensité, le mouvement, le nombre et le rythme, sont
apprehendées directement comme des attributs perceptifs globaux et amodaux." (Stern, 1985, trad.frçse p.77)

Ensuite, les affects de vitalité, s'ils ne sont pas "catégorisables" parce qu'amodaux, trouvent leur consistance dans leur "profil d'activation", leur schéma temporel interne. Franèse, en 1958, indiquait que l'expressivité musicale repose sur l'évocation d'"abstraits sentimentaux" qui seraient "les échos et les ombres de l'expérience corporelle sedimentée" (1958, rééd. 1972, p.343). L'hypothèse est intéressante, mais d'une part, l'expressivité musicale, on le voit, ne saurait être réduite à l'évocation de la seule expérience corporelle, même si celle-ci y apparaît non négligeable, et d'autre part, le terme d'"abstrait" semble particulièrement mal choisi, laissant entendre que l'abstraction cognitive serait indispensable pour que fonctionne une "reconnaissance" ou une "identification" des schémas formels inscrits dans la musique. Au contraire, le concept d'aspect de vitalité s'inscrit dans la dynamique même de la vie affective, dans la dynamique de l'émergence de la personnalité et du lien interpersonnel, dans ce que Bergson aurait lui-même appelé la créativité de l'élan vital. L'aspect de vitalité est donc un temps lui-même émergent, un fragment de temps dans le présent qui se ressent comme une suite de tensions et de détentes plus ou moins fortes, comme une suite de variations d'intensité de la sensation. Un exemple donné par D. Stern est remarquablement lumineux : pour calmer son enfant, la mère va lui dire "Allez, allez, allez...". Elle le fait en accentuant la première syllabe, en ralentissant sur la deuxième. Mais elle peut tenter d'obtenir le même effet sans rien dire, seulement en caressant la tête de son enfant : le geste, la caresse ont alors le même profil, appuyés au début, ralenti et allégés à la fin. Ce qui est intéressant ici est que le bébé ressent les deux comportements de la même façon, et qu'il fait ainsi l'expérience d'un même affect de vitalité (1985, trad.frçse p.83) caractérisé par un profil d'activation déterminé qu'il ressent et "reconnait" immédiatement. Au fur et à mesure que s'accumulent des expériences du même genre, les affects de vitalité se regroupent en organisations très globales, dans lesquelles les perceptions, les actes et les pensees n'existent pas en tant que tels, mais qui constituent la "matrice" de leur développement dans les expériences ultérieures. Ainsi, les affects de vitalité peuvent apparaître comme les catégories (sensitives et intuitives) primitives sur lesquelles se construiront ultérieurement les émotions, les sentiments, les formes perçues et identifiées, les pensées. Leur organisation reste, dit Stern, "le domaine fondamental de la subjectivité humaine", elle est "le réservoir fondamental dans lequel on peut puiser toutes les expériences de création." (ibid°, p.95).

Deux nouvelles notions apparaissent maintenant qui nous ramènent, l'une à l'expérience du temps et à la création du temps musical, l'autre à la question si longtemps controversée de la sémiotisation du temps musical et des formes qu'il revêt dans les œuvres et les styles, en un sens à la narrativité en musique.

Tout d'abord, de tout ce qui précède, la notion de trame temporelle d'éprouvé (ou trame temporelle du ressenti) découle naturellement : elle est définie dans La constellation maternelle (1995) comme forme de représentation de l'expérience affective. Elle est donc un contour d'affectivité, la forme temporelle d'un ensemble de profils d'intensité, de rythme et de durée d'affects de vitalité, dont elle assure, pour le sujet, la cohérence dans un présent qui dure. Son émergence est pour le sujet un événement qui se produit en temps réel, à l'"intérieur" de l'expérience. Mais pour comprendre la portée de cette notion, il faut tout de suite poser la question : qu'est-ce qui se trame ? Autrement dit, qu'est-ce qui se noue, comme se noue une intrigue, un événement dont on attend un récit, c'est-à-dire dont on attend qu'on nous en donne le sens ? Autrement dit encore, si "ça se trame", quel sens ça a ?

Il y a derrière cette question une interrogation beaucoup plus générale qui regarde le temps : celui-ci n'a de sens pour nous que parce qu'il est lien, passage entre passé, présent et futur, qu'il a donc une orientation où se projette notre vie. Chez les bébés, il est clair que le temps existentiel est encore fragmenté, que l'orientation générale n'existe pas, qu'il se réduit au présent, c'est-à-dire aux sensations, aux gestes et mouvements, aux actions interpersonnelles et aux affects de vitalité qui leur sont associés. Au début - nous l'avons vu précédemment - il peut
se réduire à quelques secondes, voire parfois à moins. Mais dans ce présent si bref, comment se construit la cohérence de soi et du monde environnant, comment, sans intuition d’un horizon temporel futur, le sujet peut-il repérer la direction de ce qui se trame dans l’écoulement temporel ressenti ? Bien des philosophes ont évoqué ce qu’ils appellent l’épaisseur du présent, l’extension du présent dans le temps des ressentis subjectifs. Saint-Augustin déjà distinguait entre les unités du temps objectif et les unités du temps subjectif, qui se dilatent en fonction de l’expérience intérieure, de sorte qu’on peut parler d’un passé du présent, d’un présent du présent et d’un futur du présent : "Le temps est une distension de l’âme", le futur n’est que l’attente du futur, le passé n’est que le souvenir du passé, le présent est l’attente elle-même concentrée sur l’objet qui n’est pas encore et dans un instant ne sera déjà plus.

Ce que nous trouvons chez Saint-Augustin, c’est bien cette problématique du sens, non seulement de la directionnalité du temps et de ce que le sujet en vit, mais bien du "où ça va ?" en tant qu’intentionnalité des sensations, des perceptions, des actes, des représentations. Le présent va vers quelque chose parce que la conscience du temps est conscience de quelque chose dans le temps, conscience de quelque but à atteindre : cela aussi, Husserl l’avait vu. Le temps donc unifie, la durée constitue le lien des ressentis au-delà de leur hétérogénéité, elle est co-substantielle à l’émergence du soi. Quel est ce lien ? Ou si l’on veut encore, quelle fenêtre du temps permet-elle cette cohérence, ce sens ? Cela revient à se demander comment nous pouvons avoir des représentations temporelles. Chez le nourrisson, la capacité à repérer des cohérences dans le monde interpersonnel est liée à sa capacité à ressentir le lien entre ses actions et le plaisir ou le déploiement qu’il en éprouve. Autrement dit, les actions se colorent d’une motivation intrinsèque - non posée comme but conscient, le plaisir ou le déploiement étant eux-mêmes des ressentis immédiats - qui maintient l’énergie pour agir. Tous les comportements d’auto-excitation relèvent de ce mécanisme. Sur lui, se greffe et s’accorde le comportement de la mère. Si bien que l’"éprouvé de faire" se colore d’une orientation, d’un but qui, après-coup, donne à cet éprouvé une cohérence par une limitation précise dans le temps : faire pour éprouver du plaisir, puis peu à peu pour répondre à l’autre, pour partager avec lui; mais dès que le but est atteint, l’action se recentre sur une autre séquence, la précédente étant renvoyée au passé, momentanément oubliée. C’est ce renvoi au passé qui, après-coup, donne sa cohérence à la séquence, la fait apparaître comme une forme temporelle ayant eu un commencement, un milieu et une fin.

C’est cela que D. Stern appelle l’"enveloppe proto-narrative " (1994 et 1995, pp.86-87) : en effet, la forme narrative est ce qui, dans l’univers du langage et des signes auquel le nourrisson aura accès plus tard, constitue l’unité de temps, découpe la réalité du devenir humain. L’après-coup temporel est donc une sémiotisation des profils d’activation des affects de vitalité, ou plus exactement, il est ce qui permet à la sémiotisation de se développer dans la durée, ce qui donne forme aux trames temporelles des ressentis, en somme il est ce qui fait que quelque chose "se trame", "prend sens" dans le temps. L’enveloppe proto-narrative est donc un contour d’affectivité réparti dans le temps avec la cohérence d’une quasi-intrigue : elle s’organise autour de la mise en acte d’une intention-motivation (orientation vers un but), elle découpe une portion de temps dans laquelle le bébé ressent sa propre cohérence, c’est-à-dire rapportée à soi (sens d’un soi noyau) tout en même temps les sensations de ses besoins (par exemple, avoir faim), de ses actes (mouvements, cris...), de ses perceptions (visage, caresses, voix de la mère...), de ses ressentis (affects de vitalité liés à la fois à ses sensations, ses mouvements, ses perceptions), mais elle reste, en-deçà de tout langage, une ligne de tension dramatique intuitive.

Elle est donc une forme proto-sémiotique de l’expérience intérieure du temps, une matrice du "récit" des tensions et des détériorations liées à l’"intrigue" (ou la "proto-intrigue", 1994) de la recherche d’une satisfaction, elle est ce qui donne à l’expérience son unité globale, quel qu’en soit le degré de complexité. Cette unité est "éprouvée" à travers la conscience progressive que le bébé acquiert de la causalité et de la progression de l’expérience vers son état final. Vers 3 ou 4 mois, le bébé devient en effet capable de différencier le soi de l’autre, le soi étant ressenti comme auteur ou cause d’actions délibérées. Il y a donc perception intuitive d’un soi-agent, d’une action, d’un instrument et d’un but. Ce sont ces éléments qui sous-tendent les formes primitives de la causalité et organisent des proto-intrigues, font émerger de la trame temporelle...
du ressenti une ligne de tension dramatique qui l'oriente : "Lorsque la motivation (désir) est activée en une situation interpersonnelle, elle crée subjectivement une structure quasi-narrative" (1994).

On rejoint là ce que dit P. Ricoeur dans *Temps et récit* (1983) lorsqu'il parle de "structure pré-narrative de l'expérience". Nos expériences, dans le dynamisme de ce que Bergson aurait appelé l’élan vital, sont ressenties comme des trames d'histoires qu'il faudrait peu de choses pour raconter et dont la cohérence se nouerait dans l'acte même de la narration. Peu de choses ? Une mise en code dans le langage, une distance à la réalité créée par sa représentation, une temporalité maîtrisée et surtout quelqu'un à qui raconter, fusses nous-mêmes.

Ces faits et ces réflexions apportent évidemment un éclairage nouveau au concept de *macro-structure* de la pièce musicale que j'ai développé à plusieurs reprises (1981, pp.88-90; 1991). J'ai défini la macro-structure comme un schéma de structuration du temps, c'est-à-dire comme le schéma de progression des structures temporelles de tension et de détentes de la pièce musicale. Autrement dit, une pièce musicale est d'abord une mise en ordre d'événements sonores dans le temps, la macro-structure un schéma simplifié type, une mise en ordre a priori que viennent remplir ensuite les événements sonores concrets dont la progression pour l'auditeur peut se définir comme une suite structurée et hiérarchisée de tensions et de détentes. La macro-structure se définit à la fois au plan de la grammaire musicale, des opérations cognitives en jeu dans la composition et la compréhension de l'œuvre, mais aussi au plan de l'expressivité et des ressentis de l'auditeur au cours de l'écoute. La progression temporelle, à travers les tensions et les détentes, à travers les patterns formels qui évoquent ce que je peux bien maintenant appeler des affects de vitalité, prend sens dans cette continuité orientée du début à la fin de l'œuvre, trouve sa cohérence dans cette trame temporelle qui noue des gestes mélodiques, rythmiques, harmoniques, égraine les "proto-narrations" des mille nuances de la vie intérieure dans un profil unique. Commentant le texte de l'*Essai sur les données immédiates de la conscience* de Bergson, V. Jankélévitch décrit bien cela : "Mais quel moyen terme relie jamais une douleur et une joie ? Pourtant la durée fait ce miracle." La durée intérieure, comme d'ailleurs la modulation harmonique, implique "l'intuition d'une certaine épaisseur d'originalités à franchir." (1959, pp.4-43).

Ainsi, la macro-structure apparaît clairement comme la *trame temporelle des ressentis* ou *éprouvés* organisés, affects de vitalité dessinés en leurs profils proto-narratifs, véhiculés dans le flux du temps musical. La musique serait ainsi comme la traduction ou la représentation de la matrice originelle de toutes les formes symboliques, de toutes les formes de langage, c'est-à-dire de toutes les formes de mise en ordre du temps. Au fond, on comprendrait, du point de vue de la psychologie du développement et de la psychanalyse, l'intuition profonde de Lévi-Strauss dans *L'Homme Nu* : "Mythe codé en sons au lieu de mots, l'œuvre musicale fournit une grille de déchiffrement, une matrice de rapports qui filtre et organise l'expérience vécue, se substitue à elle et procure l'illusion bienfaisante que des contradictions peuvent être surmontées et des difficultés résolues." (1971, p.590).

La musique serait-elle alors aussi la matrice originelle de toutes les formes de récits ? J.J. Nattiez a raison de poser la question au fond essentielle : l'idée de narration appliquée à l'œuvre musicale n'est-elle pas le témoin de ce désarroi de l'auditeur contemporain devant des œuvres qui n'ont plus rien de linéaire ? Donc des œuvres où "ça ne se trame pas", où il ne se passe rien, ou plutôt où rien ne passe ? Je rejoins ici les deux hypothèses les plus pertinentes d'un travail de C. Deliège (1989) : d'une part, à partir de Debussy, l'organisation interne des séquences qui composent une pièce musicale est plus importante que le processus discursif qui les lient dans la forme d'ensemble (et ceci correspond parfaitement à ce qu'a été la recherche de Debussy pour rompre toute symétrie, pour instaurer en musique la primauté de l'instant par rapport au processus qui dure, recherche que les compositeurs contemporains ont prolongée jusqu'à descendre à la séquence-son); mais d'autre part, l'exigence de l'auditeur de trouver
dans l’œuvre qu’on lui propose un minimum de clarté et d’ordre intelligible, apparaît contradictoire avec l’affaiblissement, voire la disparition du discursif. En termes de psychologie cognitive, cette exigence est en fait la possibilité de construire des anticipations et des rétrocCTIONS perceptives sans lesquelles aucune unité n’est perçue ni conceptualisable, ni en musique, ni dans quelque autre domaine de l’activité humaine de connaissance que ce soit.

Or tout une part de la musique contemporaine, en particulier la musique sérielle, privilégie les petites séquences unitaires de temps, les profils proto-narratifs élémentaires, ces vecteurs dynamiques dont nous avons parlé précédemment au détriment de la progression d’ensemble, du geste temporel portant l’œuvre dans son intégralité. La musique sérielle est construite sur ces aventures temporelles minimales qui restent sans lien. Car dans le même temps, le principe de non-répétition sur lequel Schoenberg fonde son système, rend improbable que la trame temporelle englobant ces séquences minimales définies autour de la série puisse jamais prendre sens pour l’auditeur. La “grande forme” reste purement abstraite, elle ne lie pas les enveloppes proto-narratives entre elles dans une re-présentation d’expérience vécue, elle reste étrangère à la trame de la durée intérieure et de sa possible symbolisation dans le temps de l’œuvre tout entière.

Parce qu’au fond cette musique évite la répétition, elle engendre ce temps lisse dont a parlé Boulez (1963), ce temps de l’instant, ce temps qui ne passe pas, ce temps qui ne peut plus être le temps d’une aventure et que l’auditeur ressent parfois comme le temps de l’ennui. Le refus de toute expressivité, le refus de toute sémantisation de la musique dans les années du serialisme triomphant n’est donc psychologiquement pas un hasard : il est lié à la destruction de la "grande forme", c’est-à-dire à la destruction de toute pensée du temps, de toute pensée d’une trajectoire définie anticipable dans le temps : l’aboutissement ultime de cette tendance de l’art musical de la première moitié du XXᵉ siècle est sans doute la Momentform, c’est-à-dire une forme qui n’a plus rien de linéaire et qui est entièrement centrée sur le matériel. L’absence de tout processus discursif la rend alors totalement imprévisible et indiscernable. Et, si des trois Viennois, Webern nous paraît le plus énigmatique comme le plus subtilement novateur, c’est bien que, si chez lui la variation l’emporte sans cesse sur la répétition, il a pourtant réussi un équilibre exceptionnel entre le discursif-minute et l’œuvre-sequence !

"De même, quand elle lisait la prose de George Sand..., attentive à bannir de sa voix toute petitesse, toute affectation qui eût pu empêcher le flot puissant d’y être reçu, elle fournissait toute la tendresse naturelle, toute l’ample douceur qu’elles réclamaient à ces phrases qui semblaient écrites pour sa voix et qui pour ainsi dire tenaient tout entières dans le registre de sa sensibilité. Elle retrouvait pour les attaquer dans le ton qu’il faut, l’accent cordial qui leur préexistait et les dicta, mais que les mots n’indiquent pas; grâce à lui elle amortissait au passage toute crudité dans les temps des verbes, donnait à l’imparfait et au passé défini la douceur qu’il y a dans la bonté, la mélancolie qu’il y a dans la tendresse, dirigeait la phrase qui finissait vers celle qui allait commencer, tantôt pressant, tantôt ralentissant la marche des syllabes pour les faire entrer, quoique leurs quantités fussent différentes, dans un rythme uniforme, elle insufflait à cette prose si commune une sorte de vie sentimentale et continue."

Je termine par le texte de Proust : il me fournit, je crois, une conclusion lumineuse : on sait que c’est à travers cet épisode et son contexte affectif fort (le coucher qui sépare l’enfant de sa mère, la crise de larmes et la lecture de François le Champi", "un vrai roman") que Proust entre en littérature. Le mouvement du texte est au fond l’inverse de la démarche de tous ceux qui, à propos de l’œuvre musicale, veulent à tout prix lui appliquer les catégories et les théories de la narration. Peu à peu, ce qui devient essentiel, c’est tout ce qui n’est pas le récit, si mystérieux qu’il n’éveille guère l’attention du jeune Marcel pour ce qui est de son contenu propre, c’est l’élan de la phrase, c’est le rythme de la voix, sa musique si particulière qu’elle va constituer l’essence inconsciente de la phrase proustienne, c’est la trame temporelle de la lecture, non la logique des événements. Mme Proust recrée pour son fils, et à son insu, l’enveloppe proto-narrative du roman, c’est-à-dire ce qui est en-deçà du roman, ce qui est
A VANT le roman, l'émotion particulière qui conditionne que ce récit-là (mais c'aurait pu être un autre) aura cet impact sur la formation de la personnalité du jeune adolescent. Mais c'est aussi ce qui lui permet, dans cette expérience-là, de construire sa cohérence interne, de lui son passé à Combray avec sa vocation d'écrivain.

De même, dans l'analyse des structures sémiologiques de la musique, avons-nous d'abord à retrouver ce mouvement, cette dynamique de la linéarité temporelle, cette "intrigue" sans personnages ni événements, ces fractures et ces restaurations de la cohérence de notre devenir. On ne peut parler de narrativité en musique, car la musique ne raconte pas, ne représente pas, ne dit pas. Pas plus que ne dit et ne raconte la voix de Mme Proust à son fils. Elle met seulement

en scène le temps et nos façons de le ressentir, de le reconstruire, de le fantasmer entre continuité et discontinuité, entre unité fusionnelle et morcellement, entre mobilité et immobilité. L'histoire de la création musicale est l'histoire des rapports de l'homme au temps, tant au plan individuel qu'au plan collectif. Mais le parcours à travers les remarquables travaux de D. Stern nous suggère un nouvel approfondissement : la référence constante au fait que la psychologie individuelle du temps se forge dans les interactions avec autrui, la constatation majeure que le sentiment de la durée naît dans le jeu de la communication interactive, nous font comprendre que la musique ne nous touche qu'à travers l'autre. Certes, je suis tout à fait d'accord avec J.J. Nattiez lorsqu'il montre que la "communication musicale" (au sens banal) n'est qu'une illusion. Mais je crois que le problème de l'expression musicale est ailleurs : la musique puise son pouvoir dans sa nature profondément sociale, tout comme d'ailleurs le langage en tant que véhicule de la représentation intériorisée. Toute la substance temporelle de la musique est nourrie de nos façons d'être au monde, c'est-à-dire d'être dans notre temps, notre culture, avec notre perception, notre corps, nos émotions et nos sentiments. Elle n'est pas communication, mais elle est représentation de notre pouvoir de communiquer, elle est jeu stylisé de notre ouverture au monde, elle est communication sans objet à communiquer. Elle est proto-narration, c'est-à-dire au sens propre antérieure à la narration, irréductible à ses catégories et ses décompositions logiques du réel.

BIBLIOGRAPHIE

Standing Together Under One Umbrella:
A Multidisciplinary and Interdisciplinary View of Music Psychology

Donald A. Hodges
Institute for Music Research
University of Texas at San Antonio

In many ways this is a rich and fertile period in the history and development of the field of music psychology. Several new books (e.g., Deliège and Sloboda 1996, Hodges 1996, Kemp 1996, and Pratt and Spintge, 1966) have been published in the last year. Significant research is being shared through conferences such as this one and through journal articles. The quality of research is also improving. So there is much to be pleased about current circumstances. On the other hand, there are also some concerns. Among those that might be enumerated, the one I wish to focus on is what I perceive to be a lack of agreement on what is meant by the term “music psychology.” For all its current richness, music psychology is a fractionated discipline. Much high quality work is going on here and there around the world, but often without an overarching conception of the field as a whole.

One of the reasons this is so is that differences in training may keep us from a more coherent and complete view of the field. For example, psychologists frequently have different perspectives than musicians; both may be considering the same musical behavior, but each from an orientation reflecting strengths and weaknesses arising from profiles of training and experience. Simplistically, musicians may regard some psychologists’ research as musically naive, while psychologists may view the research of some musicians as less than rigorous. Other differences arise between basic and applied researchers, between researchers and practitioners, between differences based on geography or language, and so on. As another example, consider that American scholars are often not proficient in other languages. Since a number of excellent sources are not available in English, many of us remain ignorant of a significant body of research and our view is therefore more narrow.

Of course, there can be advantages when one’s views are challenged and perhaps strengthened by previously unfamiliar ideas. We have much to learn from one another. What would be more helpful, however, is movement from a splintered to a more holistic view. Two contrasting images are offered as an analogy. In one image, there are a number of individuals, each standing under an umbrella. In the second image, the same number of individuals are all standing together under one very large umbrella labeled "Music Psychology." Standing together under one umbrella does not rob anyone of his or her individual personality. However, it does imply that different personalities are seen as parts of a more holistic world view (or Weltanschauung).

In sketching out my conception of a holistic view of music psychology, I will begin by drawing some observations from a very brief look at the historical development of the field of music psychology. Then, I will give brief synopses of contributions from philosophy, physics, biology, anthropology, sociology, psychology, and music, followed by some implications of a holistic view.
Historical Conceptions of Music Psychology

While there are certainly pioneers, as early as Pythagoras in the 6th century B.C. and later Helmholtz (1863) and others in the 19th century, Carl Seashore (1919, 1938, and 1947) is considered by most as the father of modern music psychology. Interestingly enough, Seashore (1938) believed that the proper study of musical behavior encompassed physics, physiology, psychology, anthropology, philosophy, and metaphysics. He said that while the whole field might be called the "science of music," psychologists tend to take over the field for want of a sponsor and so the field has come to be dominated by psychologists and thus called music psychology.

E. Thayer Gaston, who might be properly credited as the father of modern music therapy, supported Seashore's contention. He said that in seeking to understand musical behavior, we would be best served by looking first to the behavioral and other sciences (Gaston, 1968a and 1968b). Following the lead of Seashore and Gaston, Charles Eagle, a student of Gaston, created an interactive model for his bibliographic database of music psychology literature (see Figure 1).

![Descriptive molecule model of the interdisciplinary world of Music Psychology](after Eagle, 1996a).

The important lesson to draw from this brief historical perspective, is the multidisciplinary and interdisciplinary nature of music psychology as conceived by some of the pioneering leaders. Unfortunately, this viewpoint is not fully agreed upon, as an examination of the contents of music psychology texts will reveal.
Accepting Seashore’s rationale and following Eagle’s model, I am proposing that the term “music psychology” should mean an all-encompassing approach to the study of musical behavior that is both multidisciplinary and interdisciplinary in nature. Thus, understanding musical behavior in its totality requires a multidisciplinary view incorporating input from many different disciplines. Likewise, an interdisciplinary view is critical—integrating ideas from many disciplines, often in new hybrids such as biochemistry or social psychology. For the remainder of this paper, subsequent uses of the term “holistic view” imply this overarching, umbrella approach to the study of musical behavior.

Synopses of Contributing Disciplines

The intent of this section is to provide very brief glimpses into ways the various disciplines contribute to an understanding of musical behavior. In each case, a selection of related topics will be used to demonstrate not only "vertical" movement within a discipline, but "horizontal" movement among disciplines as well.

Contributions from Philosophy

Those who are surprised to find philosophy included in this holistic view of music psychology might be interested to read this statement: "No metaphysics, however deep, no theory of aesthetics, however firm its philosophical formulation, can discuss the musical experience and ignore psychological points of view" (Révész, 1954). Philosophical inquiry into musical behavior has most often been concerned with the meaning of music, the understanding of beauty, and our emotional responses to it. Révész’s statement and the excellent extant contributions notwithstanding, this has not been one of the more fruitful areas in the field of music psychology. Explicating musical behavior could benefit considerably from more philosophizing, particularly from those who are broadly and deeply read in the “science of music.” Although theories are different from philosophies, it seems appropriate to remark at this juncture that the field is also woefully short of coherent, guiding theories. The best research is that which is theory-driven and all-too-often experiments are not carried out within a conceptual framework.

Books such as In Search of Beauty in Music: A Scientific Approach to Musical Esthetics (Seashore, 1947), Aesthetics and Psychobiology (Berlyne, 1971), and Music and Mind: Philosophical Essays on the Cognition and Meaning of Music (Fiske, 1990) illustrate philosophical contributions to the understanding of musical behavior.

Footnote: For these descriptions I have borrowed heavily from Eagle (1996a) and Hodges and Haack (1996). In addition to the extensive citations from these sources, readers may wish to perform searches on the Computer-Assisted Information Retrieval Service System for Music (CAIRSS). CAIRSS in an online bibliographic database of more than 16,000 research articles in the fields of music psychology, music medicine, music therapy, and music education (Eagle and Hodges, 1992). It can be accessed on the world-wide web at http://imr.utsa.edu. In the CAIRSS Thesaurus, Eagle (1996b) has identified relevant articles published in over 2,230 different journal titles, in 23 languages and from 54 countries.
Contributions from Physics

Historically, the connections between music and physics are among the most ancient, beginning at least as early as the Pythagorean experiments. From that time, through the Middle Ages, when music was placed in the upper quadrivium of the seven liberal arts along with arithmetic, geometry, and astronomy, and continuing up to today’s advancements in digital sampling and MIDI technology, music and physics have shared a close association.

Acousticians study sound, the fundamental nature of music. From acoustics, further investigations lead to psychoacoustics (psychophysics). Practical applications of acoustics are made in architecture and engineering, and the physics of musical instruments leads to ergonomics and biomechanics. Researching certain problems can involve mathematics, astronomy, and quantum physics.

Introduction to the Physics And Psychophysics of Music (Roederer, 1975), The Quantum Reality of Music (Eagle, 1991), and The Music of the Spheres: Music, Science, and the Natural Order of the Universe (James, 1993) are three illustrative readings from physics.

Contributions from Biology

From biologists we learn that all human beings are biologically equipped to be musical and that there is a synergistic, symbiotic relationship between music and the body. Music behavior is shaped and constrained by biology (e.g., limitations on the vocal system so that we can only sing so high or low, or so loud; limitations on the hearing system so that we can perceive pitches only so high or low; limitations of only ten fingers with which to play the piano; etc.).

Conversely, musical experiences have profound effects on the body. Understanding these relationships requires investigations into such biomusical topics as vision, hearing, the brain, emotions, motor mechanisms, and physiological responses (including heart and pulse rate, electrodermal responses, respiration rate, blood pressure, muscular tension, blood volume, skin temperature, gastric motility, pupillary reflex, blood oxygen, hormone secretion, and pupillary reflex). Animal studies are also informative. Practical applications are made in music medicine, performing arts medicine, and music therapy.

These three books provide examples of contributions from biology: Music, Mind, and Brain: The Neuropsychology of Music (Clynes, 1982), Biomusicology: Neuropsychological, Neuropsychological and Evolutionary Perspectives on the Origins and Purposes of Music (Wallin, 1991), and MusicMedicine II (Pratt and Spintge, 1996).

Contributions from Anthropology

Anthropologists have a message for us that can be stated rather simply but which is profound in its impact on our understanding of the significance of music: All people in all times and in all places have engaged in musical behaviors. Elaborating on this theme takes us around the world and back in time. Archaeologists search for evidence of ancient musical
practices and ethnomusicologists seek to document the role of music in all the world's cultures.

As did biology, anthropology teaches us about two-way relationships. Blacking (1973) describes this eloquently in a pair of opposing chapters: "Humanly Organized Sound" (i.e., we create the music) and "Soundly Organized Humanity" (i.e., the music, in turn, shapes us). While all human cultures engage in musical behaviors, the variety of ways in which they do so is nothing short of staggering.

Contributions from anthropology may be represented by *The Anthropology of Music* (Merriam, 1964), *How Musical is Man?* (Blacking, 1973), and *The Study of Ethnomusicology* (Nettl, 1983).

**Contributions from Sociology**

While anthropologists tend to look across cultures, sociologists tend to look within a given culture. Confirming biological and anthropological findings, sociologists tell us that each individual has the potential to respond to music of the surrounding culture. No condition of age, race, gender, mental or physical state, or socioeconomic status prohibits one from a meaningful musical experience. The sociology of music takes us into business and economics, politics, religion, the military, youth culture, and the entertainment and media industries.

Most of us live in sound-saturated societies. As Merriam (1964) noted:

> The importance of music, as judged by the sheer ubiquity of its presence, is enormous.... There is probably no other human cultural activity which is so all-pervasive and which reaches into, shapes, and often controls so much of human behavior (p. 218).

Anyone who doubts the validity of this statement is invited to try this experiment: Go for 24 hours without hearing any music and note the departures from the usual routine that are necessary.

Representatives of sociology's contributions include *Sing a Song of Social Significance* (Denisoff, 1983), *Music and Its Social Meanings* (Ballantine, 1984), and *The Sociology of Music* (Dasilva, Blasi, and Dees, 1984).

**Contributions from Psychology**

There are an immense number of topics that can be explored under the rubric of psychology. Psychologists from each of the major approaches (e.g., Freudian, behavioral, Gestalt, developmental, cognitive, and humanistic) contribute unique understandings of musical behavior. A significant body of music research literature falls within each of these psychological orientations.

Psychologists are interested in the perception and cognition of music. Gardner's (1983) ground-breaking theory of multiple intelligences is causing more researchers to look at the role music plays as a human knowledge system. Other important topics include the musical personality, special musicians (e.g., musical savants, Williams Syndrome musicians, etc.), the
development of musicality, stress in performance, affective responses to music, musical aptitude, and music teaching and learning. Practical applications, particularly in music therapy and music education, are rich in their own research traditions.


**Contributions from Music**

Musicians, of course, have made important contributions to the understanding of musical behavior. In many cases, musicians have been active in other disciplines. The composer Bartok was, for example, one of the first ethnomusicologists, working to study and preserve the folk music of Hungary. From a research standpoint, three groups have made the strongest and most consistent contributions: music theorists, music educators, and music therapists.

Music theorists have particular insights into the structure of music. *A Generative Theory of Tonal Music* (Lerdahl and Jackendoff, 1983) is a significant representative. Music educators are particularly concerned with developing musical skills, as evidenced in *Handbook of Research in Music Teaching and Learning* (Colwell, 1992). Music therapists use music while working with clients with a variety of handicapping conditions. *Music Therapy for Handicapped Children* (Lathom and Eagle, 1982) provides an overview of research literature for younger clients.

Although stated at the outset, I cannot resist emphasizing once more the necessity for musical validity in research purporting to explain musical behavior. Some research studies may provide information about processing sound, but not necessarily about processing music.

**Implications of a Holistic View of Music Psychology**

What are the implications of such a holistic view of music psychology? Here are five issues to consider:

1. **Establishing Research Teams.** One of the best ways to foster multi- and interdisciplinary research is to create research teams consisting of specialists from various disciplines. Though it can take some time for individuals to learn a common vocabulary, understand each other's biases, and so on, there are enormous advantages to be gained when the collective expertise of such a group is focused on a specific research problem.

   An excellent example of this can be seen in a series of articles written by a group from the University of Rochester (Chuang et al., 1988; Crummer et al., 1988; Crummer et al., 1994; Frisina, Walton, and Crummer, 1988; Hantz et al., 1992; Swartz et al., 1992; Walton et al., 1988). This research team had participants from the schools of psychology, medicine, education, and music (including music performance, music theory, and music education).
In my own case, I am working with Peter Fox, editor of the Human Brain Mapping Journal and an expert in PET scan research, and Larry Parsons, a cognitive neuroscientist. The three of us, supported by many others, including lab technicians and statisticians, have completed one brain imaging study of pianists (Fox et al., 1997) and are currently working on a PET study of music cognition. Since my professional training is as a musician, there are limitations on the neuromusical research I can conduct on my own. Forming this research team allows us to attack a problem with the combined strength of our various backgrounds and resources.

2. Setting Research Agendas. Music psychologists ought to consider adopting a model from other disciplines in the setting of research agendas. In medical research, for example, there are frequent meetings that bring together recognized experts in a given area. These individuals meet to devise an agenda that guides research efforts for the next period of time. Documents arising from these gatherings are useful in setting priorities and in helping researchers to focus their work. (See, for example, Today's Opportunities, Tomorrow's Health: The Future of Biomedical Research in America; Swazey and Watkins, 1992.) Recognizing some of the difficulties involved, one can still envision gatherings designed to set research agendas for various topics in music psychology. Another idea is for conferences such as this one to set aside several hours each day for targeted discussions. It would take considerable effort to produce useful guides, but over time we would get better at it. Also, this would provide occasions for representatives from various viewpoints to work together toward common goals.

3. Publishing Holistic Works. Standard textbooks vary widely in breadth or narrowness of scope. Books that are more tightly focused on a specific topic also vary, either ignoring or including input from various disciplines. An excellent model of inclusion is the recently-published Musical Beginnings: Origins and Development of Musical Competence (Deliège and Sloboda, 1996). First, it should be noted that the book is published in both English and French. Second, even a brief perusal of the contents reveals a broad, multi-pronged approach. The main theme of this book is the development of musical competence as an underpinning to an understanding of the cognitive basis for musical behavior. This puts it squarely in the literature of cognitive psychology, an idea that is reinforced by discussions of such topics as language acquisition, cognitive development, communications, and temporal order. But physics is represented, for example, in discussions of the acoustical environment of the fetus in utero and in the acoustical parameters of speech and music. Biological topics include physiological development of the fetus and infant, evolution, and animal studies. Anthropological concerns are satisfied by cross-cultural comparisons of German, British, Nigerian, American, and Chinese subjects. Sociology is represented by topics such as human culture, social interactions, mother-infant interactions, and creative play. Two isolated sentences, admittedly taken out of context, further support the holistic vision of this book. "The study of musical development has grown dramatically in recent years, and this has occurred within a number of disciplines including psychology, cognitive science, education, sociology, and anthropology, not to mention music and musicology themselves" (Hargreaves, 1996, p. 145). "This means that the strong hierarchic nature of tonal music is at the same time a property of that music, culturally determined, and a psychologically and biologically
determined requirement" (Imberty, 1996, p. 200). It is to be hoped that many more publications will follow this excellent model.

4. Developing Multidisciplinary Curricula. One of the most difficult issues hindering a holistic view of music psychology is the training received by those who consider themselves music psychologists. While it is clear that no single individual can be fully conversant in all the disciplines, more emphasis should be placed on acquiring a broad background. Here is a brief outline for a Ph.D. in music psychology that we are developing at the Institute for Music Research: This 90-hour program includes a music core (18 hours of music history, theory, and performance), a music psychology core (21 hours of psychology of music, anthropology of music, biology of music, physics of music, music technology, and current issues in music psychology), research core (9 hours), cognate areas (21-hour specializations in one of the following: music perception and cognition, neuromusical research, music medicine, music for the cognitively impaired, or music technology), electives (12 hours), and the dissertation (9 hours).

5. Changing Individual Perspectives. Finally, one of the simplest and quickest changes to make would be for everyone working in the field to adopt a multi- and interdisciplinary view of music psychology. That is, we should all be invited to stand under the same umbrella. We should be encouraged to read broadly and to adopt a holistic attitude toward musical behavior. Even when we are focused on a very specific problem, as those engaged in research must be, the design of a study must take into consideration a broad array of previously published information and perhaps even more importantly, the final conclusions to be drawn from a particular experiment must not be allowed to stand alone. How these new conclusions fit into the larger picture must always be explicated.

For those of you already under our umbrella, I am pleased to be a colleague. For those of you not yet under our umbrella, we invite you to join us. Together we can do much to promote the understanding of musical behavior.

References


42
Symposium: Tristan cor anglais tune
This paper treats the solo English horn melody from the beginning of Act III of Wagner's Tristan und Isolde, first to a prolongational analysis along the lines presented in GTTM (Lerdahl and Jackendoff 1983), then to a schematic phrasal analysis that bears some resemblance to Nattiez's (1975) paradigmatic method. The melody is called the "alte Weise" [old tune] by Tristan. In the course of Act III it serves as a backdrop and lament for much of the long monologue in which he excavates his psyche and its motivations.

A prolongational analysis assigns a hierarchy to a sequence of pitch events, establishing connections that emphasize the return of structurally stable events. Within these larger connections, relatively unstable events elaborate the stable events. The overall structure can be described in terms of nested waves of tension and relaxation. A prolongational analysis derives from a combination of stability conditions (formalized in the pitch-space theory of Lerdahl [1988]) and rhythmic function (formalized in GTTM's time-span reduction). Here we will be concerned only with results of the derivation.

The prolongational analysis of bars 1-4, the first half of the first phrase of the melody, shows C prolonged by upper-neighbor motion, followed by stepwise descent to Ab. In scale degrees the motion is ^1-^5-^6-^5-^4-^3 in F minor. This progression is a diatonic elaboration of the chromatic Sehnsucht leitmotive that begins the Act I Prelude, traversing ^1-^6-^5-(#4)-^4-[#3] in A major/minor (the third scale degree is implied by the unresolved dominant seventh). This connection is set up in the Act III Prelude, which begins with a diatonic variant of the other inverted half of the Sehnsuchtmotive from the Prelude to Act I, with equivalent prolongational structures. (Inversional relationships permeate Tristan, even in the "alte Weise" at bars 14-15.) If in Act I the motive embodies unsated longing, in Act III it expresses the other side of desire, the remorse and despair of the mortally wounded Tristan. This difference is reflected in the details of melodic attractions (Lerdahl 1996): in the Act I Prelude the attraction of F (^6) to E (^5) passes through the equally strong attractions of D# to D and D (^4) to the unconsummated C# (^3); in the "alte Weise" the parallel attraction of Db (^6) to C (^5) is weakened by the interpolated G, and Bb (^4) resolves relatively weakly to a melancholy Ab (^3).

In the prolongational analysis of bars 5-9, the second half of the first phrase, the melodic bass descends to an implied C while the soprano descends to a half-cadential V^2. The most striking moment in this phrase is its final Gb. Nattiez (1990) says that the "alte Weise" evokes the lugubrious gondola tunes that Wagner heard in Venice as he was composing Act III. From my limited acquaintance with gondola tunes, I suspect that at least for a 19th-century German they characteristically emphasized b^2; see, for example, Mendelssohn's Lied ohne Wörte, op. 30/6. Stylistic references aside, this Gb also has consequences internal to the "alte Weise." On the one hand, b^2 induces a strong attraction to ^1, to which it resolves at the beginning of the next phrase; this recalls the Sehnsucht motive's original attractional force. On the other hand, the Gb acts as a discordant b^5 to the implied dominant harmony.
The second phrase (bars 10-21) covers the same prolongational territory as the first. The third phrase picks up on the upper-neighbor motion in bars 2-3 (b^7-b^6-a^5) and elaborates it with the motive from bars 5-7, again coming to partial rest on an extended V^2. The fourth phrase, after a short bridge in bar 37 (quoted from bar 24), resumes the inverted form of the melodic head from bars 14-15 and achieves full closure an octave below the melody's basic register.

At a global prolongational level, phrases 1 and 2 both move from i^5 to V^2; phrase 3 prolongs V^2; phrase 4, raised to the basic register, achieves closure with V^2-i^1. Within each phrase there is a clear linear descent of the kind favored in Schenkerian analysis. Perhaps this is Wagner's way of saying that it is indeed an "alte Weise". Formally, the phrases group into a Wagnerian Bar: A (phrase one) and its variant A' (phrase two), followed by an extended, developmental B (phrase three). The only departure from this schema is the addition of the truncated, closural fourth phrase. (This formal description corresponds to Lorenz's [1924-33] classic account.)

This brings us to a schematic analysis of each phrase of the "alte Weise," starting with an approach suggested by the Formenlehre tradition (adumbrated in this century by Schoenberg 1967, Ratz 1951, Caplin 1985, and others). In the modified approach taken here, rhythmic groups (in GTTM's sense) are labeled according to motivic and/or structural function, and a schematic form is posited in which surface regularities are "normalized" as in Schachter (1980) and Rothstein (1989). The functional categories employed in the present analysis are: basic idea, contrasting idea, and closing idea or closure; extension, contraction, and expansion; repetition and sequence; and inversion. These categories should be self-evident except for the distinction between extension and expansion; here I follow Rothstein in calling an enlargement an "extension" at the end of a phrase and an "expansion" in the middle of a phrase. Normalized phrases are assumed to be in multiples of two in bar-length, in phase with a regular underlying hypermeter. Thus in the first phrase, which takes nine bars at the musical surface, the first four bars present the basic idea (which could be broken down further); this is answered by a contrasting group, comprising the presentation and sequencing of the contrasting idea (bars 5-6) and a slightly extended closure (bars 7-9). In the normalized version, the closing Gb is deleted, yielding two 4-bar groups.

The second phrase is more complicated. In its surface form, the basic idea expands into a grouping overlap at bar 14, at which point an inversion of the basic idea substitutes for the contrasting idea; and at bar 17 the half-cadential V^2 stretches out for three bars through repetitions, contractions, and an extension. If the expansions, repetitions, and extension are removed, the result again is a schematic form of 4 + 4 bars. The developmental third phrase begins at bar 22 with the contrasting idea, sequenced in 2-bar groupings. After the arrival of V^2 at bar 26 there occur two large extensions, first to bar 31 and then to bar 36, accomplished by internal contractions and repetitions. The repetitions on the Sehnsucht motive (Gb-F-b) create an obsessive climax for the "alte Weise." If these repetitions are eliminated, joining bar 29 to bar 36, the schematic form again becomes 4 + 4 bars. After the bridge in bar 37, the final phrase returns to the inverted basic idea. The removal of the prolongational expansion in bar 40 reduces bars 38-42 to a normative 4-bar phrase. The brevity of this phrase compensates for the long exertions of the previous one; the effect is one of tragic finality.

It is useful to cast these functionally-labeled surface phrase structures in a format resembling that of Ruwet's (1966) and Nattiez's (1975) distributional method (Labussière [1992] provides such an analysis of the "alte Weise"). In the representation adopted here, the surface sequence reads from right to left on each stave, beneath which appear the
labeled groupings; and the groupings are aligned vertically so that nonadjacent parallel passages are placed beneath one another down the page. My purpose in adopting this layout is to elucidate the parallelisms that influence the assigned grouping structure: parallel passages align vertically.

The issue of parallelism is where the distributional method and the GTTM theory touch base, and in a way that is complementary. However, the distributional method and GTTM's grouping component are both seriously incomplete in that they do not state rules that generate parallel structures. Developing such rules becomes problematic once the parallelisms in question lie beyond the reach of a few surface transformations. It would be difficult, for instance, for an automatic system to identify the underlying relationship between the Sehnsuchtmotive and the "alte Weise," and yet reject other equally complex cases that do not behave as parallel structures.

But if it is beyond this paper to propose rules for parallelism, progress is feasible in a related area. The paradigmatic approach has not treated underlying levels of structure and needs to be supplemented by reductional theory. Often paradigmatic substitutions depend more on prolongational than on surface similarity. If one can construct distributional graphs for musical surfaces, one can make them for underlying levels of structure as well. This can be done for phrases in their schematic form, or, more tellingly, for global prolongational structures. As noted above, the four phrases traverse very similar prolongational structures. In these convergences at underlying levels resides the overall coherence of the "alte Weise."

References

WAGNER, Tristan's "Alte Weise": Two perceptual approaches

Irène DELIEGE
Centre de Recherches Musicales de Wallonie
University of Liège, Belgium

This study, conducted at the request of J.J. Nattiez, continues a research programme which I developed between 1986 and 1990 with the aim of elaborating a model of the mental representation of musical works in real time listening (Deliege, 1991). This model is based on the central hypothesis that the listener abstracts cues during listening, i.e. salient features provided by the emergent properties of the musical surface. By means of literal or varied reiteration, these cues take on the role of labels beneath which longer sequences are subsumed; they enable musical structures to be memorised, identified through listening, and classified in terms of two basic principles which underpin the listening process: namely, the principles of SAME and DIFFERENT.

A series of experimental procedures, developed progressively, allowed different aspects of the hypothesis to be examined (for more details see Deliege, in press; Deliege & Mélen 1997). Two of these, the segmentation procedure and the mental line procedure, which had previously been applied to pieces of contemporary repertoire, the Sequenza VI for solo alto by Luciano Berio and Eclat for 15 instruments by Pierre Boulez (cf. Deliege, 1989, 1993), are used again here. The segmentation procedure shows that, under the effect of the SAME principle, structures form groups for as long as the same basic cue is perceived. The perception of a contrast, i.e. the intervention of a new cue, interrupts this grouping process. A boundary is established by means of the DIFFERENT principle and a new group develops afterwards, centred upon the new cue which is evident to the listener. By this process the large structural articulations of the work are delineated. The so-called mental line procedure enables, at a different time, the role of these cues to be shown in the elaboration of the work in the listener's memory. Segments of the piece are presented to listeners in random orders, with the task of reconstituting the piece by positioning the segments from memory. New techniques have considerably refined the original procedure (see below). Furthermore, the current study includes the factor of training: it compares the performance of musician and non-musician subjects with different levels of musical training. Moreover, emphasis is placed here upon the factor of familiarisation: the two tasks of segmentation and reconstitution are preceded by different numbers of prior hearings. From earlier research, it is anticipated that the factors of subjects' training and number of hearings will not be significant for segmentation. However, these factors should become evident in memorisation of musical structure and the task of reconstituting the piece.

1. Segmentation

Method. Four categories of adult subjects (median age, ± 25 years) were selected in order to examine the factor of training: 22 non-musician university students (NMS), 20 musician students (MS), 22 university researchers (UR), all non-musicians, and 12 professional musicians (PM).

The piece was recorded by Ivan Dudal, soloist of the Orchestra of the Royal Opera La Monnaie in Brussels, with a duration of 2 minutes 22 seconds. A CD recording was used to play the experimental materials via the CD-ROM of a Macintosh PowerPC, with an amplifier connected to the computer by MIDI interface and two speakers placed 4 metres from each other and from the listener. The MAX program was used for data collection.

With the exception of the PM group, the other subject groups were divided into two subgroups who heard the piece either once or three times before the segmentation tasks for investigating the factor of familiarisation. The PMs did not undergo the three prior hearings. There are thus 7 sub-groups of subjects in total (NMS1, NMS3; MS1, MS3; UR1, UR3; and PM).

All the subjects were tested individually. After a familiarisation stage, two segmentation tasks were set: a simple segmentation and a hierarchical segmentation. The second task also provided a control of the stability of subjects' responses. During the familiarisation stage, subjects were instructed to listen to the piece as if it were a text within which they should introduce segmentations similar to punctuation marks in an ordinary text.
For simple segmentation, responses were made by pressing the 1 key on the numeric keypad, regardless of the weight of the perceived segmentation. For hierarchical segmentation, three different degrees were included—weak, average, strong—by pressing keys 1 (=weak), 1 and 2 together (=average) or 1, 2 and 3 together (=strong). The weighting of the segmentations and timing information (seconds and milliseconds) were recorded by the computer during the task (table 1).

### Table 1. Example of time data for a single subject's responses

<table>
<thead>
<tr>
<th>Part 0</th>
<th>Part 1</th>
<th>Part 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The table shows a list of data. Beneath &quot;part 0&quot;, there is no response.</td>
<td>1 14722</td>
<td>2 14691</td>
</tr>
<tr>
<td>Familiarisation hearing, there is no response.</td>
<td>1 29893</td>
<td>2 229948</td>
</tr>
<tr>
<td>Beneath &quot;part 1&quot;, the segmentation task, the first number &quot;1&quot; indicates that the subject was pushing a single key for segmentation. The following numbers show exact time data: the last three numbers of each line represent milliseconds, the preceding numbers seconds.</td>
<td>1 49630</td>
<td>3 494949</td>
</tr>
<tr>
<td>Number 1, the subject gave to the segmentation, and the time is as for part 1.</td>
<td>1 70000</td>
<td>2 69500</td>
</tr>
<tr>
<td>Note: For subjects who underwent 3 familiarisation hearings, their first segmentation responses appear under &quot;part 3&quot; and &quot;part 4&quot;.</td>
<td>1 118000</td>
<td>1 139412</td>
</tr>
<tr>
<td>Familiarisation hearings, their first segmentation responses appear under &quot;part 3&quot; and &quot;part 4&quot;.</td>
<td>1 118400</td>
<td>3 139542</td>
</tr>
</tbody>
</table>

**Results.** A maximum of 21 segmentations was observed (see figure 1: I to XXI). Figure 2 illustrates the segmentations made during the first hearing (black lines) and during the second (dotted lines); confirmations (grey lines) show segmentations made at the same places by the same subjects during the two segmentation tasks. Analysis of variance by the factors of training and number of familiarisation hearings shows non-significant tendencies for training \(F(3,69)=3.3, p=0.09\) and for familiarisation hearings \(F(1,69)=3.6, p=0.06\) and an absence of interaction \(F(2,69)=1.2, p=0.31\). This situation remains constant when the PM data are removed from the analysis \(F\) (training 1,58)= 3.52, p=0.07, \(F\) (familiarisation hearings 1,58)= 2.94, p=0.06 and \(F\) (interaction 2,58)= 1.18, p=0.32\).

Comparison of the two tasks in terms of segmentation stability shows that the factor of training is significant \(F(3,69)=2.83, p=0.04\), but the factor of familiarisation hearings is not \(F(1,69)=1.6, p=0.21\) and there is no interaction between the two \(F(2,69)=0.99, p=0.38\). However, if the data of the PM group are excluded, all the effects become non-significant \(F\) (training, 3,58)= 0.07, p=0.94\), \(F\) (familiarisation hearings, 1,58)= 1.36, p=0.25, \(F\) (interaction 2,58)= 0.84, p=0.44. It thus appears that the behaviour of the PM subjects differs from the other groups.

Finally, evaluation of the weighting (weak, average or strong) of the segmentations (the second task) reveals the principal segmentations made (PS). These are located at the points I, II, III, VIII, XI, XV, XVIII and XXI (fig. 1, italic roman numerals). No effect of training \(F(3,69)=0.60, p=0.62\), nor of familiarisation hearings \(F(1,69)=0.11, p=0.74\) was observed, nor was there a significant interaction between the two \(F(2,69)=0.5, p=0.83\). These results remain unchanged when the data from the PM group are excluded from the analysis.

### 2. Reconstruction

**Method.** For this task, the mental line procedure was employed. The factors of training and number of previous familiarisation hearings were also considered. Four groups of subjects (24 NMS, 20 MS, 23 UR and 12 PM) participated in this experiment.

The piece was divided into 7 segments of different length, ending at the points I, II, III, IX, XI, XVIII, XXI (see fig. 1). The musical materials were played via a MIDI interface and MAX software using 7 keys of a device called the ScaleGame, constructed by the technical staff of
Figure 1. Wagner, *Tristan’s Alte Weise*. The roman numerals indicate the 21 segmentions observed (Principal segmentations (PS) are in italics). Examples of the exact time (seconds and milliseconds) of each note inception are given on the two first staffs of the score.
Compilation of the segmentations perceived by the subjects

Figure 2. The figure shows for the 7 subgroups of subjects, the segmentations made during the 1st hearing (black lines) and during the 2nd hearing (dotted lines); confirmations (grey lines) are segmentations made at the same places by the same subjects during the two segmentations tasks.
the Centre de Recherches Musicales de Wallonie, which enabled real time listening of the content of each key. The NMS, MS and UR groups were divided into two sub-groups depending on the number of familiarisation hearings (3 or 5) they received before the task. The PM group received only 3 familiarisation hearings. The 7 sub-groups of subjects were distributed as follows: 12 NMS3, 12 NMS5, 11 MS3, 9 MS5, 12 UR3, 11 UR5, 12 PM3.

All the subjects were tested individually. The 7 keys were presented in a different random order for each subject, and subjects were instructed that after either 3 or 5 familiarisation hearings they would be asked to rebuild the piece. They were allowed to take notes during the familiarisation stage. There was no time limit for the task: timing information for each subject was recorded by the computer software.

**Results.** Out of a total of 79 subjects, 33 correctly rebuilt the piece (42%). Excluding data for the PM group, 25 of the 67 remaining subjects (37%) had correct results. This percentage was significantly influenced by training. Correct reconstruction was achieved by 5 subjects out of 24 in the NMS group (21%); by 12 subjects out of 20 in the MS group (60%); by 8 subjects out of 23 in the UR group (35%); and by 8 subjects out of 12 in the PM group (67%) ($\chi^2 = 10.58$, 3 df, p=0.01).

The number of familiarisation hearings also influenced success rate. Excluding the PM data, 8 subjects out of 35 (23%) correctly reordered the segments after 3 familiarisation hearings, compared with a success rate of 17 subjects out of 32 (53%) after 5 familiarisation hearings (p=0.02). These results are modified when data for the PM group is included: 16 subjects out of 47 produced correct results after 3 hearings, i.e. 34% against 53% (p=0.15).

Timing information did not have a significant relationship to correct or incorrect reconstruction. The mean duration for correct reconstructions was 12.9 minutes compared with 15.5 minutes for incorrect ones (Student's t, p=0.08).

The distance from the correct solution, which is defined as the sum of absolute differences between the position attributed to the segment by the subject and its actual position (cf. table 2), shows an effect of training [$F (3,72)= 5.3$, p=0.002] and number of familiarisation hearings [$F (1,72)= 9.5$, p=0.003]. There is no interaction between these two factors [$F (2,72)= 1.6$, p=0.21]. These results do not change when the data for the PM group is excluded [effect of training: $F (3,72)= 8.9$, p=0.004; effect of familiarisation hearings: $F (1,72)= 3.3$, p=0.04; interaction: $F (2,72)= 1.5$, p=0.24].

The positions attributed by the subjects to the seven segments also show an influence of training and number of familiarisation hearings. Table 3 illustrates, at the level of primary modes, that after 3 hearings a majority of only the MS and PM groups attained the correct order, but after 5 hearings the other subjects achieved correct results also.

It can also be seen (table 4) that the location of the segment in the piece is an important factor: the initial and terminal segments (1 and 7) are more accurately positioned than those in the middle.

**Conclusion.** The segmentation approach has here replicated results already observed in the studies mentioned in the introduction, suggesting that the performance of such kind of task does not appear to be dependent on musical training. Different levels of training, in terms of both general education (undergraduate students compared with university researchers) and musical education (music school students compared with professional musicians) have been tested in the experiment, but no significant variation was observed on this criterion. The new factor of familiarisation was considered here, which does not appear to have significantly influenced the task. Finally, the hierarchisation of segmentation points indicates that the segmentations which are selected as strong appear in all the responses. These results provide further support for the suggestion that cue abstraction in real time listening may be a psychological process of an automatic nature, within which operation on the basis of the principles of SAME and DIFFERENT appears to be privileged.

---

1 For information, contact the author of this paper
Table 2: Distance between subjects’ versions and the correct version

The table shows results for non-musician students (NMS), musician students (MS), university researchers (UR), professional musicians (PM), and for all subjects globally, respectively for each box, the median of the distance (1st line), the mean and standard deviation (2nd line) and the range (3rd line), after three or five familiarisation hearings, as well as globally.

The number of subjects in each group is also given (4th line).

<table>
<thead>
<tr>
<th>Segm</th>
<th>Mode</th>
<th>Primary/Secondary</th>
<th>Mean</th>
<th>Mean Deviation</th>
<th>NMS</th>
<th>3 familiarisation hearings</th>
<th>5 familiarisation hearings</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMS</td>
<td>1</td>
<td>1/2</td>
<td>2.08</td>
<td>1.08</td>
<td>9</td>
<td>8.5 (4.3)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3-5/4-6</td>
<td>4.33</td>
<td>2.33</td>
<td>2-16</td>
<td>0-10</td>
<td>n=12</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2/1</td>
<td>2.08</td>
<td>1.42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4/3</td>
<td>3.75</td>
<td>1.08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3-4-5/6</td>
<td>4.17</td>
<td>1.17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>6/5</td>
<td>4.58</td>
<td>1.42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>7/-</td>
<td>7</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>1</td>
<td>1/2-6</td>
<td>1.55</td>
<td>0.55</td>
<td>4</td>
<td>4.4 (5.1)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2/5</td>
<td>3.00</td>
<td>1.00</td>
<td>2-16</td>
<td>0-7</td>
<td>n=9</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3/1-2</td>
<td>2.55</td>
<td>0.64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4/3-6</td>
<td>4.09</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5/4</td>
<td>4.64</td>
<td>0.55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>6/3</td>
<td>5.18</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>7/-</td>
<td>7</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UR</td>
<td>1</td>
<td>1/-</td>
<td></td>
<td>0</td>
<td>6</td>
<td>5.0 (3.1)</td>
<td>4.2 (5.6)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>6/3</td>
<td>4.33</td>
<td>2.33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2/3</td>
<td>2.58</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4/3</td>
<td>3.42</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4-5/3</td>
<td>4.25</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>6/2-3</td>
<td>5.42</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>7/-</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM</td>
<td>1</td>
<td>1/-</td>
<td>1.00</td>
<td>0</td>
<td>6</td>
<td>2.0 (3.2)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>6/5</td>
<td>5.67</td>
<td>0.33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>7/-</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Parameters of the distribution - primary modes (in bold) and secondary modes, mean and mean deviation of the positions attributed to each segment by the different groups of subjects, as well as globally.

<table>
<thead>
<tr>
<th>Segm</th>
<th>Mode</th>
<th>Primary/secondary</th>
<th>Mean</th>
<th>Mean Deviation</th>
<th>NMS</th>
<th>3 familiarisation hearings</th>
<th>5 familiarisation hearings</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMS</td>
<td>1</td>
<td>1/2</td>
<td>2.08</td>
<td>1.08</td>
<td>9</td>
<td>8.5 (4.3)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3-5/4-6</td>
<td>4.33</td>
<td>2.33</td>
<td>2-16</td>
<td>0-10</td>
<td>n=12</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2/1</td>
<td>2.08</td>
<td>1.42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4/3</td>
<td>3.75</td>
<td>1.08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3-4-5/6</td>
<td>4.17</td>
<td>1.17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>6/5</td>
<td>4.58</td>
<td>1.42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>7/-</td>
<td>7</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>1</td>
<td>1/2-6</td>
<td>1.55</td>
<td>0.55</td>
<td>4</td>
<td>4.4 (5.1)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2/5</td>
<td>3.00</td>
<td>1.00</td>
<td>2-16</td>
<td>0-7</td>
<td>n=9</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3/1-2</td>
<td>2.55</td>
<td>0.64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4/3-6</td>
<td>4.09</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5/4</td>
<td>4.64</td>
<td>0.55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>6/3</td>
<td>5.18</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>7/-</td>
<td>7</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UR</td>
<td>1</td>
<td>1/-</td>
<td></td>
<td>0</td>
<td>6</td>
<td>5.0 (3.1)</td>
<td>4.2 (5.6)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>6/3</td>
<td>4.33</td>
<td>2.33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2/3</td>
<td>2.58</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4/3</td>
<td>3.42</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4-5/3</td>
<td>4.25</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>6/2-3</td>
<td>5.42</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>7/-</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM</td>
<td>1</td>
<td>1/-</td>
<td>1.00</td>
<td>0</td>
<td>6</td>
<td>2.0 (3.2)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>6/5</td>
<td>5.67</td>
<td>0.33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>7/-</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4. Positions attributed to the seven segments by subjects

For each of the 7 possible positions (in bold), the numbers in the columns indicate which segment was placed in that location, as well as the number of subjects (in parentheses) who attributed this position to that segment. Values for Student’s t and probabilities were calculated on the global data, means and mean deviations (see table 3).

<table>
<thead>
<tr>
<th>segm</th>
<th>1 (67)</th>
<th>2 (5)</th>
<th>3 (11)</th>
<th>4 (12)</th>
<th>5 (1)</th>
<th>6 (9)</th>
<th>7 (9)</th>
<th>t value</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>1 (9)</td>
<td>2 (23)</td>
<td>2 (5)</td>
<td>2 (2)</td>
<td>2 (3)</td>
<td>-</td>
<td>-</td>
<td>3.07</td>
</tr>
<tr>
<td>2</td>
<td>2 (41)</td>
<td>2 (23)</td>
<td>2 (13)</td>
<td>3 (9)</td>
<td>3 (6)</td>
<td>-</td>
<td>7.22</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3 (11)</td>
<td>3 (38)</td>
<td>3 (13)</td>
<td>4 (17)</td>
<td>4 (7)</td>
<td>-</td>
<td>-1.12</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5 (14)</td>
<td>5 (4)</td>
<td>5 (9)</td>
<td>5 (43)</td>
<td>5 (7)</td>
<td>-</td>
<td>-4.34</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6 (8)</td>
<td>6 (5)</td>
<td>6 (7)</td>
<td>6 (56)</td>
<td>-</td>
<td>7 (79)</td>
<td>-8.97</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>(*)</td>
<td>(*)</td>
<td></td>
</tr>
</tbody>
</table>

Reconstitution was examined by means of the procedure known as the Mentaline. However, this procedure has been greatly altered in the sense of presenting the totality of the musical material to subjects. Thereby this represents an intermediary form between the Puzzle paradigm ² (Deliege et al., 1996) and the pure form of the Mentaline. In earlier research (Deliege, 1989, 1993), only certain extracts of the piece were to be positioned by subjects. In contrast to the segmentation approach, the factors of training and familiarisation play an important role here, as illustrated in the way that cues are memorised. This finding relates to those of earlier studies (Deliege, 1989, 1992, 1993), in particular concerning the effects of primacy and recency: the internal segments of the piece receive more tentative positioning than those at the start and the end.

Finally, the particular composition of segment 2 reveals a clear effect of the principles of SAME and DIFFERENT in reconstituting the piece. This segment is located between segments 1 and 3, whose musical content is initially identical. A large number of subjects were confused by this segment 2, and on many occasions the segments 1 and 3 were placed next to each other. Furthermore, this segment 2 had an effect on the total task duration: even if the subjects finally achieved the correct solution, this segment caused significant time loss. This may be the reason behind the lack of any significant differences between observed durations for correct and incorrect solutions.³

some references


² It should be emphasised however that the only borrowings from the Puzzle paradigm here are the division of the piece into segments and the notion of using the musical material in its entirety. The psychological processes under investigation by these two procedures differ as follows: the Puzzle, in asking subjects to reconstruct a piece from segments which they have not heard, taps the declarative processes of the individual, whilst the Mental Line procedure investigates the mental schema which is evoked during listening.
³ I am grateful to Alexandra Lamont for help in translation.
In 1992, while I was working on an analysis of the Tristan's cor anglais solo, I invited Irène Deliège and Fred Lerdahl to write their own analyses. In the case of Fred Lerdahl, I wanted to be able to compare the results of his method with a paradigmatic analysis inspired by Ruwet (1972: chap. 3 and 4) and the prolongational analysis following Meyer's model presented in Explaining Music (1973). In the case of Irene Deliège, I wanted to obtain an experimental analysis which could be compared with Lerdahl's analysis and mine, because the Lerdahl and Jackendoff model (1983) and a part of my semiological approach claim to be relevant for the study of perception. Today, we have the chance to present some results of our own methods and to compare them. I want to express my gratitude to both of them for having accepted to undertake these analyses. I am also very pleased that Irene Deliège invited my colleague and friend Michel Imberty to join. Since he is handling this solo from a psychoanalytical standpoint, it will allow us during the discussion to compare our approaches with an hermeneutical one, a problem which surely is at stake in musicology today.

The analysis I am showing now, intends to illustrate aspects of the so-called semiological tripartite model I have outlined in Music and Discourse (1990). This model distinguishes between the immanent (or neutral) level of analysis which describes structural phenomena, the poietic analysis which deals with the compositional strategies, and the esthesic analysis which deals with the perceptive ones. The taxonomical and prolongational analyses are both considered as immanent (or neutral) analysis until they are interpreted poietically or esthesically. In order to propose such interpretations of the immanent structures, the musicologist needs to use a theory of compositional or perceptive strategies. This is called the inductive-poietic or the inductive esthesic analysis.

If the musicologist is using documents of poietic relevance, such as sketches, letters, statements by the composer, he/she is proposing what I call an external/poietic analysis. If he/she is using experimental devices - as Irène Deliège will do - in order to try to understand how the piece is perceived, he/she is proposing an external/esthesic analysis.

It will be highly interesting to try to define the pertinence of Michel Imberty's approach according to this way of classifying the various analytical situations.

Let first look at a paradigmatic analysis of this solo (because of the time, it doesn't illustrate all the possible paradigmatic relationships in this piece). I will select those aspects which allow me to demonstrate how they can be interpreted poietically and esthesically.

The paradigmatic analysis shows on line 1 a theme which is repeated and transformed on lines 3 and 5. It reappears on line 23: the rhythm is the same but the melody is inverted. The chart shows also what are the main musical ideas which are developed: the bar 3 of the theme (on line 1), the phrase II on line 2; a new mode-like idea appears at the end of line 14 which leads to phrases VI and VII.

These remarks belong to an immanent or neutral analysis, since so far they are not poietically or esthesically interpreted.
Let first consider the relationship between lines 1 and 3 on one side, and line 23 on the other. It is obviously poetically relevant, because of what we know of the compositional strategies of the Western composers. The inversion of a melody is a very common compositional device since the Ars Nova, particularly in the contrapuntal practice. But is it esthetically relevant? In other words, once the listener reaches the hearing of line 23, does he establish cognitively a relationship between lines 1 and 3, and line 23? This is precisely a question the experimental analysis might tell us. If it doesn't, we have to use the inductive esthetic analysis, that is to interpret this structural relationship from the standpoint of a perceptive theory. Leonard Meyer provides one in Explaining Music when he deals with the perception of what he calls «conformant relationships»: «The greater the variety of intervening events and the greater the separation in time between two comparable events, the more patent the shape of the model must be in a conformant relationship is to be perceived. Or, to put the matter the other way around: the more regular and individual the pattern (and, of course, the more alike events are in interval, rhythm, etc.), the greater can be the temporal separation between model and variant and the greater the variety of intervening motives, with the conformant relationship still recognizable.» (Meyer 1973: 49) In our particular case, the variety of intervening events and the distance between the two events surely don't help for the perception of the similarity of patterning, which even is stronger for the rhythmic aspect than for the melodic one. I doubt this relationship can be perceived.

Methodologically, the immanent description allowed us to establish a relationship. This relationship seems to be relevant as far as the compositional strategies are concerned, but not from the perceptive standpoint.

The paradigmatic chart shows an other interesting phenomenon. Lines 1 and 2 demonstrate that the phrase 2 is prolonging with G and G flat a descending line initiated at the end of phrase I (C-B flat-A flat). The napolitean G flat will be completed by the tonic G with the very last note of the solo. The G and the G flat of line 2 are becoming polar notes around which the yodeling of lines 14 to 21 (or phrases VI and VII) is organized. The yodel figure G-C-C of lines 14 to 17 is obviously derived from the end of line 9, and it could be shown paradigmatically that this interval of ascending fourth is a transformation of the initial ascending fifth F-C.

This last remark raises a question from a poietic standpoint: did Wagner establish a connexion between the ascending fifth of lines 1 and 3, and the ascending fourth at the end of line 9? This is the type of question which can be answered within the framework of the so called external poetics, in this case, by the examination of sketches. Fortunately, three sketches of this solo do exist in the Bayreuth Archives. The second one demonstrates an hesitation of Wagner between the ascending fifth and the ascending fourth: when the beginning of phrase I reoccurs, the composer writes not F-C but G-C. In the final version, he will keep this idea for the end of phrase IV (line 9).

And of course, our preceding observations raise esthetical questions. The paradigm shows a descending line C-B flat-A flat-G-G flat. The two last notes are emphasized since they become polar notes between the lines 14 and 21. Probably, the final F is perceived on line 23 as a prolongation of the ending of line 21. But is it perceived as the completion of the descending melody of lines 1 and 2? I doubt of it, at least for one of the reasons brought up by Meyer for the conformant relationships: too manythings have happened between the end of line 2 and the end of line 23.

With the exception of the prolongation of lines 1 and 2 toward the ending of line 23, this paradigmatic analysis is a taxonomic one in the sense that it leads to define musical units. Using the model of melodic analysis proposed by Meyer in Explaining Music I present here a prolongational analysis of this solo. This model allows to discover bi or trilinear melodic structures. Again, this analysis may be questioned from the standpoint of the esthetics. The model of the perception of unit I seems to fit with the perceptual reality, both concerning the relationship between the gap and the fill, and the hierarchical
relationship between the main and the secondary line. But do we really perceive this solo according to a hierarchy of two or three voices? I would be interested to know if this question can be solved with the experimental tools the cognitivists use today.

I hope to have demonstrated the usefulness of the tripartite model for cognitive studies. The so called immanent or neutral level of analysis allows to elicit non interpreted structures, both taxonomic and prolongational. Inductively, we may establish if some aspects of these structures are poietically or esthesically relevant. The knowledge of external data - sketches or listerner's reactions - may help us to decide if some of them are poietically or esthesically relevant. The comparison of the inductive esthesic analysis with the results of the experimental approach will tell us if the perceptive inductions have been right. The immanent analysis may also establish a list of questions to be solved by the experimentalists. During the discussion, I hope to know if my colleagues agree with these views.

REFERENCES


PROLONGATIONAL ANALYSIS (I)
PROLONGATIONAL ANALYSIS (II)
**IV° CONGRES INTERNATIONAL DE L'ESCOM**

**Upsala, Juin 1997**

*Quelques jalons pour une réflexion d'herméneutique psychanalytique à propos du solo de cor anglais du III° Acte de Tristan*

Prof. Michel IMBERTY  
Université de Paris X - Nanterre

Dans l'un des chapitres de son livre *Le Combat de Chronos et d'Orphée* (1993), J.J. Nattiez aborde le problème de la signification du solo de cor anglais qui suit le *Prélude du III° Acte* de Tristan. L'interprétation qu'il propose est que cette mélodie est à la fois une expression de la solitude de Tristan, solitude métaphysique de celui qui s'éveille du royaume onirique de la mort et se prépare à y retourner, solitude aussi de l'être dévoré par le désir dévorant que Schopenhauer désigne comme vouloir absurde, et en même temps, la symbolisation de la fameuse mélodie infinie, c'est-à-dire la symbolisation dans l'œuvre du processus de création wagnérien, la représentation peut-être inconsciente de toute la douleur de la création, je devrais dire de l'enfantement de l'œuvre, où nous reconnaissons que Tristan est aussi Wagner.

Dans ses conférences au Collège de France (1993) qui paraîtront prochainement et dont il m'a communiqué amicalement le texte, J.J. Nattiez relie cette double interprétation à une analyse des sources externes de ce solo, et une analyse des structures qui permet de montrer comment la musique met en scène ce rêve de R. Wagner en même temps que le travail créateur.

En m'appuyant sur certains des éléments de son travail, je voudrais proposer une autre approche herméneutique du solo et de son traitement ensuite dans tout l'acte.

*Je partirai de cette constatation : l'acte s'ouvre sur une image sonore du vide. Le contraste des deux registres extrêmes des cordes dessinent l'enveloppe d'un espace creux et béant d'où toute temporalité serait exclue, une sorte d'espace et de temps morts. Ce vide silencieux d'où émerge l'antique mélodie ne se comblera que dans la plénitude de l'immense crescendo orchestral final qui, pourtant, conduit à la dissolution du désir, à la disparition de l'être dans l'immensité infinie de l'univers nocturne de la mort.  

Ailleurs (Imberty, 1993), j'ai décrit ce paradoxe de l'*utopie de comblement* dont l'origine est à rechercher dans l'inconscient le plus archaïque, précisément celui où le temps ne vient pas encore creuser le Soi de l'absence, du vide, de la séparation et de l'attente indéfinie du retour à l'état initial de plénitude, on devrait dire de retour au plein psychique initial.

Je commencerai donc par quelques remarques sur cette figure sonore du vide, dont se lit l'écho inquiétant chez P. Valéry :

"Amère, sonore et sombre citéme,
Sonnant dans l'âme un creux toujours futur !"

Je suggererai une comparaison avec une autre œuvre musicale où ce vide initial crée une tension insoutenable : l'*Adieu du Chant de la Terre* de G. Mahler. Le début, là aussi, suggère un grand vide sans voix qui vient hanter l'auditeur. Il y a seulement ce sentiment d'un espace..."
dans lequel bruissent des résonances et des vibrations secrètes et hostiles, le sentiment que les registres extrêmes (le groupe-pédale de l’ut grave, les arabesques aiguës du hautbois et de la flûte, et, entre ces limites, la tierce mineure funèbre des cors) composent les contours d’une enveloppe sonore vide d’où émergera bientôt la voix blanche, dont on ne sait si elle est une voix qui s’éveille ou se meurt. Deuil, mais aussi attente... *Morendo.* La voix qui chante dans ce vide semble détimbrée, c’est-à-dire sans individualité, elle est voix neutre, sans passé, sans avenir. De cette neutralité, bien des mélancolies donnent l’apparence, sans angoisse, indifférentes.

Au début du 3ᵉ acte de *Tristan*, la voix est remplacée par le solo de cor anglais, chant de pâtre dira Kurwenal, mais qui est bien autre chose pour Tristan. Cependant ce vide initial est délimité dans la succession par le mouvement des contours qui le dessinent, non dans la superposition des plans sonores : le lent trait des violons vers l’aigü semble ouvrir l’horizon du temps et de l’espace, mais il l’ouvre sur le vide et sur l’absence. Temps qui s’esquisse, s’ébauche, temps qui pourtant s’immobilise dans le cercle scénique du néant (image du cercle tant de fois reprises par Wieland Wagner dans ses mises en scène). Commence alors, indifférente, aussi morne que l’horizon qui borne le regard de Tristan, l’"antique mélodie" qui n’est encore que signal convenu pour l’annonce d’une non-entrée en scène, d’un non événement : l’absence d’Isolde. Un temps hors-temps, forclos, sans changement ni mouvement, un temps amorphe, le temps mort d’un refoulement.

Tout le 3ᵉ acte sera pourtant la reconquête d’une plénitude, d’une harmonie suprême, d’une "höchste Lust". Il part donc de ce vide, de ce temps mort pour finir sur un temps plein, sur un remplissage du néant par l’être. Mais en même temps, ce plein renvoie au vide absolu, celui de la mort. Ce plein que porte la vague immense du crescendo du *Liebestodgesang* n’est donc qu’illusoire, ce n’est qu’une utopie. Combler l’espace et le temps qui séparent Tristan d’Isolde, mais aussi surmonter l’attente indéfinie du retour à l’état initial de plénitude, on devrait dire de retour au plein psychique initial. Utopie d’un paradis perdu où vie et mort ne seraient plus les termes antagonistes des destins de la pulsion, mais seulement les moments alternés au miroir d’un éternel retour, utopie où le *Vide* ne serait pas le *Néant*, mais "ouverture de l’espace de l’être" (Nef, 1969).

L’*expérience transitionnelle du temps* est à l’origine de ce paradoxe. L’expérience transitionnelle, selon Winnicott (1951), peut en effet être caractérisée comme celle d’une continuité entre le monde intérieur et le monde extérieur, entre ce qui est éprouvé comme sentiment du Soi et ce qui est éprouvé comme expérience de l’environnement et des autres. Expérience de l’entre-deux, de ce qui n’est tout à fait à l’un, ni tout à fait à l’autre, ni tout à fait ceci, ni tout à fait cela, ni tout à fait moi, ni tout à fait un autre. Il s’agit d’une phase précoce essentielle dans le développement de la personnalité, où s’amorce la différenciation entre le Soi de l’enfant et ce qu’il commence à percevoir comme un Soi autonome et distinct, celui de sa mère. Phase cruciale, douloureuse, que l’enfant surmonte grâce à l’illusion de continuité que crée la pensée paradoxale dans une aire psychique qui en est le lieu privilégié d’exercice, l’*aire transitionnelle*, peuplée d’objets et d’êtres qui ne sont ni tout à fait internes (perceptions, sensations et vécus propres à lui) ni tout à fait externes (objets et personnes distincts de lui et situés hors de lui). La reviviscence à l’état de fantasmes de cette aire transitionnelle et des expériences paradoxales qui y sont liées constitue une donnée essentielle de la vie psychique en général, et toutes les activités du sujet adulte peuvent en être marquées au cours des vicissitudes de la vie, en particulier lorsque la peur de la séparation et de la perte domine. La création artistique est souvent l’occasion d’une symbolisation des expériences transitionnelles, en même temps qu’un mode de leur dépassement.

Dans son livre *Les Mains du Dieu Vivant* (1969), Marion Milner analyse la figure du cercle d’une manière fort suggestive pour nous : le cercle, dit-elle, est moins une clôture qu’un trou. C’est la représentation d’un manque, c’est un blanc, c’est une béance. Comment sont construites les partitions de Mahler et de Wagner ? Comme une série de limites qui entourent, circonscrivent : au centre, rien. Toute l’ambiguïté de ces deux figures sonores est là, essence-même de l’utopie de comblement : un plein pourrait s’instaurer à partir d’un vide, une temporalité à partir d’une intemporalité. Mais poursuivons : toujours selon Marion Milner, le
cercle vide "pourrait être en relation avec la pulsion vers l'indéterminé, un état qui peut être à la fois ressenti comme étant tout et rien" (1969, p.313), et qui donc peut être proche d'un état de mort psychique comme d'un état de plénitude, du fait que la pulsion vers l'indéterminé peut sembler une pulsion vers l'infini, ou plutôt ne se réalise que dans une quête infinie de son propre objet qui est quelque chose comme une absence d'objet, un trou, un cercle vide, un anti-objet comme on dit de l'anti-matière. On retrouve ici cette négativité essentielle, qui n'est pas un néant, et qui constitue le pendant de la positivité manifestée dans l'aire et le temps transitionnels. Ainsi le temps mort relèverait lui aussi d'une continuité entre le monde intérieur et le monde extérieur, mais cette continuité conduirait à l'illusion du sans fin, de l'ilimité, de l'infini, et comme le dit A. Green, dans cette perspective, le temps mort serait le temps de la fascination pour l'ilimité qui rejoindrait celle pour l'éternité. "Ewig...ewig !"1; "Ertrinken, Versinken, Unbewusst, Höchste Lust !"2.

Cependant cette éternité-là n'est pas l'immortalité : on conçoit ici que ce grand vide, ce cercle-trou, cette cavité résocratique qui semble illimitée est bien quelque chose comme une "aire transitionnelle négative", qui assure une continuité entre le vide intérieur et le vide extérieur, c'est-à-dire entre le sentiment de l'indéterminé de la pulsion dans le Soi et le vide d'une enveloppe silencieuse. C'est cette continuité "transitionnelle", au sens de Winnicott qui soutient cette organisation psychique paradoxalement du plein et du vide, et fonde dans l'inconscient l'utopie du comblement. Agonie initiale où la déliquescence des formes mélodiques et l'incertitude tonale donnent l'illusion d'une plénitude à l'état d'ébauche, la possibilité encore d'une ouverture. Mais ouverture vers quoi ? "Die Welt schläft ein", ou bien "Unbewusst"... Sommeil illimité de la mort ?

Tentons maintenant de saisir le mécanisme sous-jacent à cette utopie du comblement. Voici que, dans ce vide initial, ou que nous prenons seulement pour initial, l'image d'un mouvement de remplissage se profile : dans l'Adieu, le soleil s'efface derrière les montagnes, l'ombre nocturne gagne les vallées, l'espace mort, le temps mort s'animent d'une agonie. Dans Tristan, après un dialogue convenu entre le père et Kurwenal, dialogue dont la fonction est essentielle scénique, le temps se meut par l'éveil de Tristan. Éveil qui est éveil à la mort autant qu'à la vie, éveil qui est éveil d'agonie. Tel est le premier élément de notre paradoxe : l'agonie est en effet, dans la musique, une image de mouvement, de durée, une représentation d'une réalité temporelle. Elle achève, elle conduit à la mort.

Or dans les deux exemples qui nous occupent, l'agonie est en position paradoxalement rapport à ce qu'elle signifie : dans le temps mort et l'espace vide, elle devient signe d'un commencement, mais d'un commencement qui finit puisqu'il se fait morendo (indication portée par Mahler juste avant le n°4 de la partition), puisqu'il se fait appel de mort dans Tristan, désir d'oubli de ce qui commence là, le long développement de l'antique mélodie. Commencement d'un mouvement qui se construit sur le vide, ne peut exister que par ce vide, comblement de l'espace-temps mort pour retourner en fin de compte à la mort.

Dans un article fort intéressant ici, Winnicott (1974) a décrit deux ensembles de troubles pathologiques liés à ce qu'il appelle "la crainte de l'effondrement" et aux phénomènes transitionnels. Le premier ensemble est décrit sous le terme d'"agonies primitives" : il s'agit d'angoisses primitives très violentes qui concernent la peur d'un échec de l'intégration psychique, de la perte du sens de la réalité, de la chute, de la faillite de la "résidence" dans le corps (sentiment que le corps n'est pas attaché au noyau du Soi, que le corps n'est pas mien, n'est pas partie de mon identité). Toutes ces angoisses ont donc trait à la peur de ne pouvoir construire des liens entre les parties du moi, entre le moi et les autres, entre le moi et le monde. Ces agonies primitives ont donc trait à la mort psychique et à la destruction du sujet, c'est-à-dire à la crainte de l'effondrement, et elles développent des organisations défensives de la personnalité qui en bloquent le développement normal. Fantasme d'effondrement et fantasme du vide, le lien est immédiat, et les défenses peuvent s'organiser en ce que j'appelle l'utopie de comblement.

1 Dernières paroles de l'Adieu.
2 Dernières paroles d'Isolde à la fin de l'opéra.
Le second ensemble de troubles pathologiques est lié à ce que je pourrais appeler la forclusion émotionnelle de l'événement qui a provoqué la crainte d'effondrement, ou, si l'on veut, la "néantisation" de l'événement passé. Winnicott en décrit simplement le fonctionnement : "La crainte d'effondrement peut être une crainte d'un événement passé dont l'expérience n'a pas encore été éprouvée. La nécessité d'éprouver cette expérience est équivalente au besoin de se remémorer dans l'analyse des psychonévrosés. On peut appliquer cette hypothèse à d'autres craintes apparentées : j'ai mentionné la crainte de la mort et la quête du vide." Voici qui suggère une autre compréhension du temps vide, du temps mort : temps où il ne se passe rien et où, en même temps, il s'est passé quelque chose. Temps qui sans cesse se remplit et se vide du seul fait que pour le sujet, ces événements qui le remplissent ne son pas éprouvés. Ils sont des événements blancs dont le sujet va sans cesse rechercher la réalité et l'éprouvé. Il nous faudra tenter de découvrir cet événement passé non vécu, et qui tente d'émerger dans le délire de Tristan.

Je vais maintenant m'appuyer sur quelques données du travail de J.J. Nattiez ainsi que sur quelques remarques concernant le réemploi du matériel sonore du solo de cor anglais dans la scène 1 de l'acte 3, pour dégager, à partir de cette problématique d'ensemble, le sens plus profond de l'antique mélodie.

Dans ses conférences au Collège de France, J.J. Nattiez montre bien toute la richesse culturelle codifiée dans la mélodie et sa structure. Il constate d'abord, comme je viens de le rappeler à plusieurs reprises, que la mélodie triste du père a une signification très dramatique pour Tristan, signification qui échappe à Kurwenal. Elle est résurgence du passé dans la conscience de Tristan, elle est une "reviviscence", quelque chose qui ressemble à ce qui est revécu douloureusement du trauma par le patient au cours de la cure analytique. "Die alte Weise..." : alte, ancienne, lointaine, perdue dans le passé. Elle interpelle Tristan, elle lui fait rechercher ce qui est enfoncé, elle l'oblige à tenter de renouer les fils de son destin. "Qui suis-je? D'où est-ce que je viens ? Où suis-je ?" Je vais revenir sur cet aspect qui me semble essentiel et renoue avec l'ensemble de ce qui précède. La mélodie a donc un rôle de déclencheur du travail sur la mémoire et sur l'inconscient.

Cependant J.J. Nattiez note qu'elle introduit aussi - au moins du point de vue de l'auditeur du XIXème siècle - un fort sentiment de nostalgie liée à son origine populaire, le jodel que Wagner connaît au moins par ses séjours en Suisse. Mais ce sentiment de nostalgie n'est pas seulement lié au caractère "romantique" de ce qu'évoque le contour mélodique repris au jodel : il est plus directement sentiment de nostalgie du pays (natal) perdu. Nattiez en fait une belle démonstration sur laquelle je ne reviens pas, tant elle est convaincante. Or, comme il le dit aussi, le pays perdu de Tristan, ce n'est pas la Cornouaille où il a laissé Isolde, c'est le pays de la mort, le vide d'où il vient, le néant où il retourne. La mélodie du cor anglais devient ainsi expression d'une "Sehnsucht nach dem Tod", d'une aspiration à la mort qui est une des expériences qui va venir combler le vide initial que nous avons décrit. Nostalgie, mais aussi solitude qui vient cette fois du chant de gondolier entendu à Venise, dans la nuit, et que Wagner évoque dans un passage du Beethoven. Appel d'un batelier au loin qui émeut Wagner lors d'une insomnie : la solitude renforce à la fois le désespoir maléfique que suscite la nostalgie, mais ce qu'entend sans doute Wagner à ce moment-là, c'est bien une voix qui vient remplir l'espace vide, une voix qui anime le temps immobile de l'obscurité nocturne.

Si je reprends maintenant ensemble ces éléments, je peux montrer facilement que l'antique mélodie vient remplir le temps de Tristan qui s'éveille, mais le remplit en direction du passé, en direction des souvenirs refoulés. L'évocation de la mélodie enclenche le travail de recherche de ce qui est enfoui, de reconstruction des fils d'un destin que Tristan ne comprend plus. La mort des parents oubliée, la blessure, le premier voyage en barque (sur quel fleuve d'un mythisque passage des morts ?), la blessure guérie par Isolde, le philtre, et maintenant l'attente de l'accomplissement, la mort.
Un destin, c'est une histoire personnelle que je peux raconter, donc mettre en ordre. Pour cela il faut que les événements qui la composent en soient vécus (au sens émotionnel du terme), que je puisse me les approprier par le souvenir de cette vivance, qu'enfin je puisse les mettre en ordre dans le temps en les reliant les uns aux autres de manière qu'ils me soient cohérents. Or toute la scène qui précède l'arrivée de la nef ramenant Isolde en Karol est justement l'effort de Tristan pour relier les uns aux autres, dans un temps cohérent de destin, tous les événements qui ont fait sa vie. Voici donc ce comblement du temps, voici donc pourquoi le point de départ en est cette figure du vide et voici comment Tristan tente de le combler. Vie et mort, amour et mort, passé, présent, futur, un destin est un temps plein imaginé. L'antique mélodie a donc aussi une fonction de lien au sens psychanalytique, en ce qu'elle crée une unité et une cohérence dans les fragments de vie que Tristan tente de se réapproprier.

J'ai cru pouvoir noter pour ma part que 4 éléments prennent une signification psychologique assez nette au cours de cette scène. Tout d'abord, après l'intervention à des fins purement scénique de Kurwenal, c'est l'élément "jodel" qui se fait entendre en premier lieu, pivotant autour du sol bémol dans une sorte de giration atonale (quinte diminuée). Quelque chose se cherche, une durée s'ébauche, sans directionnalité particulière (p.726 de la partition Eulenburg). Ici, c'est seulement l'éveil, le retour au monde qui nous, c'est l'absence et l'attente du comblement de l'absence. Ce même élément introduit ensuite le retour de Tristan sur lui-même pour provoquer le retour de ce qui a été refoulé et lui reste encore inaccessible, forçons dans l'inconscient (p. 816). Le souvenir proprement dit et l'effort pour le faire réapparaître est très nettement associé à l'élément chromatique descendant. Tristan commence toujours à chanter sur ce motif, ou immédiatement après, et toujours pour s'interroger sur le sens de cette vieille mélodie : "Die alte Weise, was weckt sie mich ?" (p. 726); "Muss ich dich (so verstehn) ?" (p.816). Ensuite, c'est encore plus net avec l'évocation de la mort du père, " Durch Abend wehen drang sie bang, ...",(p.817-818). Plongée donc dans l'inconscient que favorise peut-être la solitude du chant de gondolier (cet élément venant de là, selon J.J. Nattiez). Tout ce qui suit semble confirmer cette analyse : p. 819-820, après l'évocation encore de la vieille mélodie sur la reprise du début du thème du cor anglais, c'est l'élément de Venise qui vient à nouveau souligner l'interrogation, le retour sur soi que suscite la mélodie, c'est l'élément "jodel" qui souligne l'évocation de la mort de la mère, plus douloureux encore que celle du père par la culpabilité qu'elle comporte ("ma mère morte par moi"), et surtout les questions fondamentales : "Zum welchem Los erkoren, ich damals wohl geboren ?". L'indétermination tonale, une fois encore symbolise l'impossibilité de mettre au jour le refoulé, le trauma. Impossibilité aussi parce que ce qui a été ainsi "oublié" appartient à ce monde de la nuit d'où l'on vient dans l'inconscience du rêve, où l'on retourne par le rêve et l'anéantissement nocturne proche de l'état de mort. La nostalgie de ce qui est perdu est bien en correspondance avec l'être "voué à la mort" qu'est Tristan.

Jusqu'ici, je n'ai rien dit du début de la mélodie. Ne serait-il répété que par pur procédé de composition ? La quinte initiale est "appel" dit J.J. Nattiez. Psychologiquement, cette interprétation me semble confirmée en ce que cet élément initial, simplement rappelé aux pages que je viens d'analyser, devient la matière du développement qui suit : ce que Tristan "découvre" dans son "passé", c'est que la mélodie s'est associée à la guérison de sa blessure par Isolde, c'est donc qu'elle est un appel à se soumettre à la loi du désir. "Désire, désirer, Mourant encore, Désire, désirer sans cesse... charmé implacable qui me contraint à vivre"). L'appel du désir c'est bien aussi à la fois l'appel du paradis perdu, et l'appel de la mort dans la solitude où le laisse l'absence d'Isolde. La quête du vide et de la mort sont l'expression de la crainte de l'effondrement. Mais il y a plus : dans son désir, Tristan en vient à maudire non point ce désir qui le dévore, mais bien le philtre dont il comprend enfin qu'il le contraint à la souffrance de l'absence, du manque, de la perte de l'objet aimé. Ce philtre, qui, au 1er acte, est absorbé par lui est un philtre d'oubli, pour Isolde c'est un philtre de mort, et l'on sait que la réalité est encore autre. L'absorption du philtre n'est pas liée à l'amour et au désir, mais pour Tristan, elle aurait dû être le début du deuil (l'oubli d'Isolde promise à Marke). Au 3e acte qui nous intéresse, le philtre est enfin revécu comme philtre d'amour qui produit le désir sans fin, le désir qui creuse le temps et l'être, le désir qui fait échec au deuil et à l'oubli.
dans lequel bruissent des résonances et des vibrations secrètes et hostiles, le sentiment que les registres extrêmes (le groupe-pédale de l'ut grave, les arabesques aigües du hautbois et de la flûte, et, entre ces limites, la tierce minore funèbre des cors) composent les contours d'une enveloppe sonore vide d'où émergera bientôt la voix blanche, dont on ne sait si elle est une voix qui s'éveille ou se meurt. Deuil, mais aussi attente... Morendo. La voix qui chante dans ce vide semble détimbrée, c'est-à-dire sans individualité, elle est voix neutre, sans passé, sans avenir. De cette neutralité, bien des mélancolies donnent l'apparence, sans angoisse, indifférentes.

Au début du 3e acte de Tristan, la voix est remplacée par le solo de cor anglais, chant de pâtre dira Kurwenal, mais qui est bien autre chose pour Tristan. Cependant ce vide initial est délimité dans la succession par le mouvement des contours qui le dessinent, non dans la superposition des plans sonores : le lent trait des violons vers l'aigü semble ouvrir l'horizon du temps et de l'espace, mais il l'ouvre sur le vide et sur l'absence. Temps qui s'esquisse, s'ébauche, temps qui pourtant s'immobilise dans le cercle scénique du néant (image du cercle tant de fois reprisées par Wieland Wagner dans ses mises en scène). Commence alors, indifférente, aussi morne que l'horizon qui borne le regard de Tristan, l'"antique mélodie" qui n'est encore que signal convenu pour l'annonce d'une non-entrée en scène, d'un non événement : l'absence d'Isolde. Un temps hors-temps, forçlos, sans changement ni mouvement, un temps amorphe, le temps mort d'un refoulement.

Tout le 3e acte sera pourtant la reconquête d'une plénitude, d'une harmonie suprême, d'une "höchste Lust". Il part donc de ce vide, de ce temps mort pour finir sur un temps plein, sur un remplissage du néant par l'être. Mais en même temps, ce plein renvoie au vide absolu, celui de la mort. Ce plein que porte la vague immense du crescendo du Liebestodgesang n'est donc qu'illusoire, ce n'est qu'une utopie. Combler l'espace et le temps qui séparent Tristan d'Isolde, mais aussi surmonter l'attente infinie du retour à l'état initial de plénitude, on devrait dire de retour au plein psychique initial. Utopie d'un paradis perdu où vie et mort ne seraient plus les termes antagonistes des destins de la pulsion, mais seulement les moments alternés au miroir d'un éternel retour, utopie où le Vide ne serait pas le Néant, mais "ouverture de l'espace de l'être" (Nef, 1969).

L'expérience transitionnelle du temps est à l'origine de ce paradoxe. L'expérience transitionnelle, selon Winnicott (1951), peut en effet être caractérisée comme celle d'une continuité entre le monde intérieur et le monde extérieur, entre ce qui est éprouvé comme sentiment du Soi et ce qui est éprouvé comme expérience de l'environnement et des autres. Expérience de l'entre-deux, de ce qui n'est tout à fait l'un, ni tout à fait l'autre, ni tout à fait ceci, ni tout à fait cela, ni tout à fait moi, ni tout à fait un autre. Il s'agit d'une phase précoce essentielle dans le développement de la personnalité, où s'amorce la différenciation entre le Soi de l'enfant et ce qu'il commence à percevoir comme un Soi autonome et distinct, celui de sa mère. Phase cruciale, douloureuse, que l'enfant surmonte grâce à l'illusion de continuité que crée la pensée paradoxale dans une aire psychique qui en est le lieu privilégié d'exercice, l'aire transitionnelle, peuplée d'objets et d'êtres qui ne sont ni tout à fait internes (perceptions, sensations et vécus propres à lui) ni tout à fait externes (objets et personnes distincts de lui et situés hors de lui). La reviviscence à l'état de fantasme de cette aire transitionnelle et des expériences paradoxales qui y sont liées constitue une donnée essentielle de la vie psychique en général, et toutes les activités du sujet adulte peuvent en être marquées au cours des vicissitudes de la vie, en particulier lorsque la peur de la séparation et de la perte domine. La création artistique est souvent l'occasion d'une symbolisation des expériences transitionnelles, en même temps qu'un mode de leur dépassement.

Dans son livre Les Mains du Dieu Vivant (1969), Marion Milner analyse la figure du cercle d'une manière fort suggestive pour nous : le cercle, dit-elle, est moins une clôture qu'un trou. C'est la représentation d'un manque, c'est un blanc, c'est une béance. Comment sont construites les partitions de Mahler et de Wagner ? Comme une série de limites qui entourent, circonscivent : au centre, rien. Toute l'ambiguité de ces deux figures sonores est l'essence-même de l'utopie de comblement : un plein pourrait s'instaurer à partir d'un vide, une temporalité à partir d'une intemporalité. Mais poursuivons : toujours selon Marion. Miller, le
Nous avons donc maintenant toutes les pièces du puzzle : le comblement du temps, la plénitude de l'être ne sont que dans le néant de la mort qui met fin au désir : c'est ce que nous dit la philosophie de Wagner-Schopenhauer. Mais si nous suivons Winnicott, ce vide ne peut être réellement comblé que si la crainte de l'effondrement est surmontée par la reviviscence d'un événement traumatique qui a eu lieu mais n'a pas été vécu émotionnellement. Tant que cette reviviscence ne se produit pas, le vide et la peur du vide demeurent, le comblement n'est qu'une illusion, une quête du plein toujours recommencée.

Or ce qui frappe dans la structure de cette mélodie, ce sont les nombreuses répétitions qu'elle comporte, répétitions de motifs brefs, variés, qui vont constituer le matériau sonore de la 1ère scène de l'acte. Tout à l'heure, chacun a montré l'importance de ces répétitions et de leurs variations. Dans les extensions qui suivent, lorsque Tristan part à la recherche ce qui est enfoui dans sa mémoire, la répétition tourne à la répétition obsessionnelle. Recommencer toujours le même acte, en revivre toujours les émotions et les sentiments qui l'affectent jusqu'à ce qu'enfin, quelque sens surgisse... Pourquoi la mélodie du père recommence-t-elle sans cesse ? Parce que le navire de l'Isolde n'arrive toujours pas ? N'accompagne-t-elle pas en réalité autre chose ? Ne veut-elle pas faire resurgir de l'inconscient de Tristan-Wagner ce qui a eu lieu et n'a pas été vécu, ou n'a pas eu lieu pour celui qui pourtant peut, rationnellement, le nommer dans le passé abstrait de l'univers ?

Cet événement, c'est l'absorption du philètre. Lorsque Tristan prend la coupe qu'il reçoit des mains d'Isolde, il ne sait pas ce que signifie le philètre, il boit seulement pour l'oubli. Brusquement, le philètre le projette dans un autre monde, sans lien avec le présent, sans lien avec tout ce qui l'entoure, sans lien imaginé avec son passé : la volonté d'oubli, le déclenchement du travail du deuil ont eu lieu, mais n'ont pas été vécus par Tristan. C'est une coupure radicale dans son histoire qu'introduit le philètre trompeur. L'horreur et la haine qu'il lui inspire dans son délire au troisième acte sont directement associées au début de la vieille mélodie, soumission à la loi du désir qu'enfin il revit. Par cette répétition de l'acte passé enfin investi affectivement, le philètre vient réinvestir de douleur et de mort toute l'histoire personnelle de Tristan, réinvestir de douleur et de mort l'absence qu'en vain il cherche à combler, la blessure en lui qu'en vain il veut refermer. L'attente indéfinie de l'Isolde est la figure de cette crainte d'effondrement qui fait aspirer sans cesse au plein-néant de l'oubli, comme la mélodie infinie est peut-être la crainte d'effondrement qui habite Wagner.

Tristan- Wagner, ai-je dit. En effet, J.J. Nattiez propose un autre niveau de signification de l'antique mélodie "qui ne meurt jamais", qui ne meurt qu'avec l'accomplissement du désir dans le grand vide de l'univers où les amants "voués à la mort" se dissolvent : elle est peut-être une représentation dans l'oeuvre de la mélodie infinie, une projection dans l'oeuvre du désir de Wagner créateur. Mais alors, cela signifie-t-il que l'amour pour Mathilde n'est lui-même qu'un événement qui a eu lieu et que Wagner n'a pas émotionnellement vécu... ou qu'il n'a vécu que dans et à travers Tristan ? Et que tout l'opéra ne serait à son tour que l'antique mélodie qui viendrait dénoter l'intrigue amoureuse, en préparer déjà chez l'auteur le travail du deuil, et le défendre de la crainte de l'effondrement ? Inévitablement, on pense à Mallarmé et à cette angoisse de la page vide qui habite

"Le poète impuissant qui maudit son génie
A travers un désert stérile de Douleurs."

Wagner est aussi de l'Azur hanté.

REFERENCES
Symposium: Analysis and Modelling of Cognitive Dynamics underlying Musical Behaviour
This paper is intended to illustrate, on the basis of experimental evidence, the general model and the theoretical frame of reference, according to which the papers presented in this session are performed.

According to our systemic-cognitivistic model, the psychological system is conceived so as to include both the organism - playing the leading role - and the environment, as inseparable interacting components. In the so-defined psychological system intrasystemic interactions are regulated by the principle of the minimum in relation to the different parameters of the intrasystemic stable states involved.

At the level of living systems, therefore, the internal-external boundary or barrier is no longer simply given but serves to transmit information into the interior of the system in such a way as to lower its entropy. In the unified viewpoint of the organism-environment system, a simple operative value is assigned to the distinction between subject and object, understood as sub-systems interacting by virtue of their essential structures. At the same time, the specification of boundaries on the basis of psychic experience makes it possible to extend modelization infinitely within the system.

Adaptation, and consequently every cognitive behavior, is therefore conceived as an intrasystemic process of re-equilibration in the single organism-environment system. Given the infinite number of systems, in the flow of re-equilibration processes the optimal form of equilibrium will be that of relative equilibrium, where each of the strategies is optimal only in relation to the strategies adopted by the counterpart. Adaptive re-equilibration, in fact, always occurs following subject’s experiences modalities both of disequilibration and of equilibrium.

In terms of architecture, the model of the mind proposed involves a dual articulation in the horizontal and vertical directions with connections distributed in a parallel fashion but hierarchically organized and the possibility of processing simultaneously in the 2 directions: bottom-up and top-down: this ensures the possibility of spontaneous modifications of the modalities of information processing as complexity increases.

The conceptual model also involves a close connection between the hierarchical-type functional organization of mental activities and an analogous organization existing in the cerebral structure. The hierarchical structure is embedded in a more general structure to be understood as the specific modality of functioning of the mental processes. This links the common hierarchical structure to “cognitive style”, personality characteristics, and attitude towards problems to be solved. The latter factors are regarded as corresponding to states of dynamic equilibrium with the external environment, regulated by specific “principles of the minimum” establishing the direction of the evolutionary processes.

In the course of our research we were able to show that these evolutionary processes are regulated by stable states present in the organism subsystem and that in cognitive processing these states are related to and influenced by subject’s spontaneous rhythms.

The first empirical evidence of an internal rhythm facilitating cognitive processing was reached by Valentini in 1951 during a research about the possibility of the contemporary performance of two antithetic tasks: a motor one, consisting in drawing going and incoming lines, and a task of silent counting. Evidence was found about the spontaneous emergence of an internal rhythm that made easier the contemporary performance of the two tasks; on the
contrary, these could not be performed when the internal rhythm did not emerge. In a following research on motor rhythm Valentini (1956) drew a difference between rhythmic strategies - spontaneous versus voluntary - adopted by the subjects; besides this he found that there is a mathematical ratio linking internal rhythm, that serves as a rule for one of the two tasks, and the automatically performed one.

The hypothesis of the existence of a spontaneous rhythm, advanced by Valentini and Fraisse, was cast in a precise structural form in the musical rhythmic model formulated by myself: In this context spontaneous rhythms could be regarded as stable cognitive states present in all subjects with a direct influence on the processing of information obtained from the environment. These particular stable states have two possible forms of structure: binary rhythm, i.e. isochronal units consisting of a strong beat and weak one, or ternary rhythm, isochronal units characterized by one strong and two weak beats.

The results of a series of studies designed in order to identify rhythmic components in a series of simple motor actions, like repeatedly pressing a computer key in tranquil environmental conditions, confirm previously acquired evidence regarding the general model, and the mathematical ratio linking internal rhythm to motor task performance. Further, the analysis of the principal components demonstrated the existence, in normal adult subjects, of one activity with a ternary scansion and another with a binary scansion, only seldom present in the same subject; in deaf adults, on the contrary, both kinds of scansions are present in all subjects simultaneously. According to our general model, we interpreted this evidence as a proof of the fact that the spontaneous rhythm specialization emerges during the cognitive interactions between organism and environment.

By the way, we directed our investigation to the detection of rhythmic structures in spontaneous musical activity. In this situation subjects are required to solve a problem of intrasystem equilibration by integrating into the processing autoctone, internal, and environmental informations. In this case, after having put into evidence the existence of the 2 rhythmic typologies, seldom present in the same individuals, the research for a connection between the level of analysis of the developmental dynamics underlying the response and the level of the cerebral dynamics shows that while the developmental dynamics of binary rhythmical structure are strongly affected by information input, ternary rhythmic structure is able, at least in finite temporal sequences, to prevail over environmental input, maintaining the deterministic character of local events when the oscillating structure falls into a chaotic situation. Analysis of the information obtained with regard to the underlying dynamics of identified rhythmic sequences made it possible to establish the presence of attractors related to the oscillating ternary structure. These attractors may be regarded as similar to those underlying the deterministic dynamics of the nervous system.

In order to illustrate the general model the next point of investigation is concerned with rhythm processing in music perception. In this case the reequilibration process is activated by the environmental input and, unlike the spontaneous internal rhythm, the pattern is multidimensional, and not only as consequence of the multidimensionality of music (Simon and Summer 1968).

Listening to whatever piece of the western musical tradition implies an implicit consciousness of two sorts of temporal organization that are mutually interdependent in musical fruition. These ones are: the “grouping structure”, according Lerdahl and Jackendoff, and the rhythmic structure that runs below the first one.

The rhythmic structure is based on the convergence of three elements: metre, speed and rhythm. The metre is the rule setting measures duration and therefore the regular periodicity of beats accents; the speed or tempo is the temporal interval among the beats; the rhythm is the dynamic articulation of metre and tempo, serving the specific musical discourse, that is the melody.
The expressive transformation distinguishing rhythm from metre appears to have particular relevance for the correct processing of a musical message, helping to reduce ambiguity as the temporal microstructure adapts to phrasing, metre and harmonic progression.

In fact, in the perception of music the listener processes the incoming temporal information by means of all available indicators in order to solve the parsing problem: metre, tempo and rhythm are, therefore, not distinguished in perception, as they converge in reducing the ambiguity of the parsing problem. Moreover, every human performance stress the convergence of the musical grouping structure with the rhythmic structure, in order to lessen the ambiguity of a message that has a development extended in time. This is the reason for which in some researches of ours we obtained a clear evidence as regard the preference for the human rhythmic performance (irregular and expressive) with respect to the perfectly regular computerized performance; this preference is more marked among women, subjects with sound musical expertise, subjects with spontaneous ternary rhythm and towards pieces with ternary rhythm. Capability in processing informations related to the peculiarity of the musical event is enhanced by sex, expertise and specific sensibility towards rhythmic structure: in fact, the analysis of subjects’ musical intelligence showed that the factors responsible for preference towards human performance are the capabilities in detecting the better rhythmic accent and the better harmonization.

These results introduce 2 new problems concerning the relations between musical appreciation and musical behavior on one side and the differential development of musical intelligence on account of sex, experience and auditory efficiency on the other.

The results that we obtained in various researches support the general model of the dynamic reequilibrations in the organism–environment system, showing a high correspondence between musical intelligence and musical behavior. Moreover experience does not improve musical attitude as it reaches complete maturation at the stage of abstract reasoning, while appreciation capacities are generally improved by experience. Finally, sex differences are partially related to age and experience; none of the 2 sexes appears to have an absolutely superior talent but each one emerges for some abilities that differ according to age and experience.

In conclusion our research activity shows that cognitive musical abilities are:

1) present in every human being (in deaf subjects too) with different grades and peculiar qualities in each individual;

2) they are strictly correlated to personal characteristics of intelligence and personality;

3) they develop along with cognitive abilities of conceptualizing and abstracting, until the stage of abstract intelligence is reached;

4) they don’t develop anymore if they are not trained before the full development of intelligence;

5) they are magnified by an instruction fit to the intelligence developmental stage and, in some aspects, they may improve, by instruction and training, also after the completion of abstract intelligence;

6) they are “transferable” abilities as they improve general cognitive information processing: in particular, evidence was reached of their influence on visual and image processing, foreign languages learning, mathematics and science learning, contemporary performance of different cognitive tasks, strategies planning and, more in general, in all cognitive tasks requiring space-temporal reasoning, attention endurance, long term memory, cognitive-motor coordination;

7) they foster the better performance in complex cognitive tasks whenever subject is simultaneously engaged in music processing.

More detailed reports about our newest research activity on rhythm processing and modelling, development and assessment of cognitive musical abilities in deaf subjects, function of long term memory in music processing, and neural modelling of musical processes will be presented in the following contributions to this session.

73
REFERENCES


INTRODUCTION
From a psychological point of view, a rhythmic structure is a configuration that emerges in time, as its single units are not simultaneously present in the perceptive field. Phenomenologically the rhythmic configuration manifests itself as an alternation of strong and weak beats. Accents, when they objectively exist, determine the specific metric interpretation, but as they may either be missing, ambiguous, or misleading, the same rhythm can be interpreted by using different metres, and the same metre can be used to interpret different rhythms (Palmer & Krumhansl, 1990; Olivetti Belardinelli, 1996). Difficulties arise when an attempt is made to consider both temporal and stress factors and to integrate objective and subjective elements. Moreover the issue is made even more complex because of possible links between the processing of rhythmic structures and other cognitive functions, such as attention and memory (Iones & Boltz, 1989), and several results support the hypotheses that these processes themselves may be rhythmically organised (Olivetti Belardinelli, 1993). For this reason modelling rhythm processing is a highly complex and difficult task, because several components, i.e. perceptive, affective, attentional, mnestic and motor ones, are involved. In this paper, in order to better understand both objective and subjective aspects of cognitive rhythm processing we decided to investigate and model entrainment mechanisms.

ENTRAINMENT MECHANISMS
An oscillation is a periodic event, characterised by a phase and by a period. The period $p$ is the length of time between the beginning of a cycle and the beginning of the next one, whereas any specific moment of the cycle is identified by the phase, usually expressed as $\phi = t/p$, where $0 < t < p$. Two oscillations are synchronised if they have the same period and the same phase at any time. Two different coupled oscillations can achieve synchronisation if one of them (driver oscillator) is able to affect the other (driven oscillator) by altering its phase or its period or both, in such a way that they regularly come into phase. The process by which two or more oscillations become synchronised is called entrainment, and it takes place by means of two mechanisms of mode-locking: phase-tracking and frequency-tracking, and both mechanisms tend to obtain the phase-locking.

The dynamics of the simplest model of oscillator is based on an activation function that increases linearly with time. This model is able to engage a phase or a frequency tracking with another oscillator and to achieve a phase-lock, but it immediately reverts to its original period, when the incoming stimulus ceases: this model is too simple for psychological modelling. For this purpose we have to assume a model with a continuous activation function, involving more complex systems of coupling, and requiring more analytical instruments for mathematical implementation.
THE MODEL

Large and Kolen (1994) proposed a model of oscillation in which a driven oscillatory unit can synchronise its continuous output to an external driver oscillator, adjusting its phase and period at certain points in time, thus reproducing an attentional pulse.

The model states that the basic oscillatory unit has a proper intrinsic period, and is able to synchronise itself with an external, periodic and discrete signal, thus if we assume \( 0 \leq t \leq T \), where \( t \) is the time and \( T \) is the total coupling duration, then \( s(t) = 1 \) at the onset of the signal, and \( s(t) = 0 \) elsewhere. Moreover the unit has a continuous and periodic output (varying from 0 to 1, and assuming a value different from 0 only in correspondence of an output pulse) determined by:

\[
o(t) = 1 + \tanh(\gamma \alpha(t))
\]

where \( \gamma \) determines the width of an output pulse (varying from 0 to 8), so that when \( \gamma \) increases the width of the pulse decreases, and vice versa, and:

\[
\alpha(t) = \cos\left(\frac{2\pi}{P}(t - t_0)\right) - 1
\]

where \( P \) is the intrinsic period of the unit, \( (t - t_0) \mod P \) is the phase.

If the signal occurs outside the output pulse, the unit does not respond to it, whereas if the signal occurs inside the output pulse, the unit tends to modify its own phase and period, so that to minimize the following error function:

\[
E(t) = s(t) \cdot (1 - o(t))
\]

This behaviour is obtained by means of a modified gradient descent procedure, by which the units tend to align its points of maximum (the peak of the pulse) with the signal pulse. The phase-tracking behaviour is obtained by minimizing the error function on \( t_0 \), because this value indirectly controls the phase

\[
\text{phase} = (t - t_0) \mod p
\]

yielding the following delta rule:

\[
\Delta t_0 = -\eta_1 \cdot s(t) \cdot \text{sech}^2 \gamma \alpha(t) \cdot \sin\left(\frac{2\pi}{P}(t - t_0)\right)
\]

The frequency-tracking behaviour is obtained by minimizing the error function on \( P \), because this value indirectly controls the period

\[
p = p_{\min} + \frac{(p_{\max} - p_{\min})}{2} \cdot (1 + \tanh P)
\]

and

\[
p_{\max} = \frac{4}{3} \cdot p_{\min}
\]

yielding the following delta rule:

\[
\Delta P = -\eta_2 \cdot s(t) \cdot \text{sech}^2 \gamma \alpha(t) \cdot \sin\left(\frac{2\pi}{P}(t - t_0)\right) \cdot \frac{\partial P}{\partial P}
\]

\( \eta_1 \) and \( \eta_2 \) are the coupling strength for phase and frequency tracking respectively. Usually there are many problems in determining the appropriate values for this two parameters. Large and Kolen (1994) proposed to use the Poincaré map, or circle map (in particular the sine circle map), to describe the dynamics for such a type of system, and by reiterating the map, to calculate the regime diagram, or Arnold’s tongues. These diagrams permit to determine the
regions of resonance for the system’s parameters, i.e. the values that make stable the system and permit a rapid convergence.

On a strictly mathematical basis the link between the model proposed by the Authors and the circle map theory is not clear. In order to understand this aspect we realised a net of oscillators, based on the circle map dynamics, and in which we have explicitly chosen the period range for each unit in a related stable region of the regime diagram (Tirozzi, Imperiali & Di Matteo, 1996).

However, the model proposed by Large and Kolen takes into account a number of aspects related to rhythm processing, in particular:

- perception, because the unit is able to modify its behaviour in presence of external signals;
- attention, because it responds to the signal only when an output pulse takes place;
- memory, as it retains its new behaviour even when the signal has ceased.

Moreover an array of oscillatory units (each one with a different intrinsic period) exhibits a complex behaviour, and indeed seems able to isolate and remember single periodic components of a rhythmic pattern.

Nevertheless, since the oscillatory units are not connected, it fails to explain how single periodic components can be integrated in a coherent perceptual configuration. Moreover the Authors say little about $\gamma$, and how to assign an appropriate value to this parameter. In the model $\gamma$ controls the degree of focused attention, and in this view it should be determined and varied by the system itself, instead of an external and fixed setting. Finally the unit is not able to evaluate its own accuracy in the response, and it does not show any explicit recognition of the signal.

In order to clarify these aspects we have reproduced the model slightly modifying the dynamics.

**CONTRIBUTION**

Firstly we reproduced Large and Kolen’s model and obtained the same results of, confirming that:

- a single oscillatory unit can synchronise itself to an external discrete and periodic signal (Di Matteo, Olivetti Belardinelli & Tirozzi, 1997),
- an array of unit can capture single periodic components of an external rhythmic pattern (we used a sample derived by means of the ‘set markers’ procedure included in Patchwork by IRCAM), and the combined output of the synchronised units yields a rhythmic configuration.

Successively, even assuming $0 \leq t \leq T$, where $t$ is the time and $T$ is the total coupling duration, we introduced the following variations to the model:

- determination of **number of unit responses** to the signal, until to the time T, by means of a unit counter

  \[
  \text{unit counter}(t) = \text{unit counter}(t - 1) + 1 \quad \text{IF } s(t) = 1 \text{ AND } o(t) > 0 \quad (9)
  \]

- determination of **number of signal occurrences**, until to the time T, by means of a signal counter

  \[
  \text{signal counter}(t) = \text{signal counter}(t - 1) + 1 \quad \text{IF } s(t) = 1 \quad (10)
  \]

- detection of the **first signal appearance**, since $t = 0$, by means of a control switch:

  \[
  \text{switch} = \text{OFF} \quad \text{IF } \text{signal counter}(t) = 0 \quad (11_a)
  \]

  \[
  \text{switch} = \text{ON} \quad \text{IF } \text{signal counter}(t) \geq 1 \quad (11_b)
  \]
determination of the average error in the response, on the basis of the error function (3) until the time \( t \), and the unit counter (9) at the time \( T \):

\[
E(T) = \sum_{\text{unit counter}(T)} E(t)
\]

- random start of the \( \gamma \) value, and a subsequent \( \gamma \) auto-setting in correspondence of the first external signal. This variation has been introduced in order to improve the entrainment behaviour of the unit. Because \( \gamma \) controls the width of the region in which \( o(t) = 0 \) [an high value for \( \gamma \) correspond to a large region of \( o(t) = 0 \), whereas a low value for \( \gamma \) correspond to a smaller region of \( o(t) = 0 \)], and the delta rules \((5, 8)\) tend to minimize the error function \((3)\) adjusting unit phase and period to the signal's owns, only when \( o(t) > 0 \), and not when \( o(t) = 0 \), we need to reduce \( \gamma \) in relation to the error value:

\[
\gamma = \begin{cases} 
\text{random value} & \text{IF switch = OFF} \\
\gamma_{\text{max}} - E(t) & \text{IF switch = ON}
\end{cases}
\]

- estimation of the ratio \( \frac{p}{q} \), between the period \( p \), finally achieved by the entrained unit and the period \( q \) of the signal. The ratio corresponds to the total number of oscillation performed by the driven unit (unit counter) divided by the total number of the external signal (signal counter). We consider the following estimate as a first approach to the real measure, and elsewhere we will analyse it in more detail:

\[
\frac{p}{q} = \frac{\text{unit counter}(T)}{\text{signal counter}(T)}
\]

In such a way the unit is able:

- to evaluate its own level of accuracy in the response \( E(T) \),
- to set its own degree of attention (\( \gamma \) variable),
- to recognise explicitly the external signal \( (p/q) \).

Successively we tested the model considering the combined output of an array of 6 oscillators (each one with its own intrinsic period) responding to an isochronous external signal; even in this case we found that the response is a rhythmic configuration.

If we consider the average error reported by each single unit, and the estimation of the rational ratio \( \frac{p}{q} \) determined by the same unit, we note that:

- the lowest average error corresponds to the unit that entrains the signal with \( \frac{p}{q} = 1/1 \)
- the subsequent \( \frac{p}{q} = 1/2 \)
- \( \frac{p}{q} = 1/3 \)
- \( \frac{p}{q} = 2/3 \)
- and so on

The progression seems to develop according to a generalised Farey Tree, a mathematical formalism that allows to generate all rational ratios, beginning from 0/1 and 1/1, characterising various levels of resonance or mode locking between two oscillations in a circle map dynamics (Treffner & Turvey, 1993).

A subset of units only achieves a stable synchronisation with the signal: this synchronisation respects the same order as the \( \frac{p}{q} \) ratios; i.e. if only one unit synchronises itself with the signal, it achieves a ratio 1/1, if two units synchronise themselves with the signal, they achieve respectively the ratios 1/1 and 1/2, and so on.
Finally only the units effectively entrained with the signal set the maximum value for $\gamma$, with a narrow output pulse, well centred on the signal; whereas the other units keep lower the same value, with a larger output pulse.

**CONCLUSION**

These results support a number of experimental evidence concerning the perception of rhythmic structures and the functioning of other related cognitive processes.

Our major result is the appearance of a rhythmic configuration in the array of oscillatory units coupled with an isochronous signal. This result is in accordance with a number of findings about subjective rhythm and phenomenal accent (Bolton, 1894; Woodrow, 1909; Valenti, 1951; Fraisse, 1978; Handel, 1984). Moreover, in previous works it was shown that experimental subjects shown tend to impose their own metre to isochronous sequences, by grouping both auditory and visual elements in sets of two or three (Olivetti Belardinelli & Besi, 1993; Di Matteo, 1995).

Our modelling trials fit with a number of cognitive models of music processing; in particular it is coherent with every hierarchical level theory of metric representations (Martin, 1972; Povel, 1981, 1984; Lerdahl & Jackendoff, 1983; Palmer & Krumhansl, 1990). According to these theories, in perceiving a rhythmic structure, subjects consider simultaneously at least 2 or more levels of beat segmentation, where the unit at one level is a multiple entire of the unit at the subsequent level. In this view we could see each oscillatory unit as the organisation of one possible level of beat segmentation.

Finally the model gives a reasonable account for some cognitive theories foreseeing links between rhythm perception and other cognitive functions, such as memory and attention. Indeed it supports the hypothesis of the existence of a temporal clock (Treisman et al. 1990, 1992) and constitutes a solid basis for the hypothesis advanced by Jones (1976, 1987) and Jones & Boltz (1989) considering memory and attention as intrinsically rhythmic processes.

**REFERENCES**


INTRODUCTION

The aim of this work is to simulate the influence of spontaneous rhythms on cognitive dynamics by means of a connectionist model. The theoretical background is derived from Valentini’s hypothesis (1970) on the existence of spontaneous rhythm and was developed by Olivetti Belardinelli in series of experimental researches showing that spontaneous rhythms, as inner stable states, influence various cognitive strategies such as discriminative learning, visual perception music processing, cognitive problem solving (Olivetti Belardinelli, 1984; Olivetti Belardinelli, 1991, 1992, 1994, 1995; Olivetti Belardinelli e Besi, 1993). According to these premises, firstly we tried to build a connectionist model of the influence of spontaneous rhythm on cognitive processing. We used a feedforward three layered network in which there was a threshold, modulated by binary or ternary rhythmic variations for the input to the units of the hidden layer. The learning and generalisation task was that of discriminating between distorted musical stimuli with vocal or instrumental performance. According to the results, recognition is better performed by networks with internal rhythm than by those without internal rhythm; further, it was found that rhythm produces a better learning but only in the presence of low amplitude, low learning speed, and enough learning time, as is in experiments on human subjects. On long time scales however, there is the possibility of an equivalence of binary and ternary rhythms, at least with regard to the performances (Olivetti Belardinelli e Pessa, 1994). Subsequently, by means of an usual error backpropagation learning rule, the same network, with and without internal rhythm, was trained to discriminate between human and computerized musical performance. In this case, besides differences in efficiency, it is possible to hypothesize an interaction between the internal rhythm of the network and the metre of the piece. We used 160 stimuli equally extracted from the human and computerized versions of the same 6 pieces already administered in the experimental sessions. Of these, 60 pieces were used for the network training; the remaining 100 stimuli, 50 for each type of rhythmic performance, were used in the generalization phase. The stimuli had been prepared by means of the Visi-pitch sound analyser. For each of them a 9 seconds period, starting with the beginning of a musical phrase, was memorized; this period was then decomposed in 32 contiguous and sequential steps; for each of these the mean frequency and intensity was calculated and normalized. Perhaps, as consequence of the stimuli step decomposition the results don’t reach statistically significant levels. However some interesting indications were reached as regard better performances: 1) in networks with internal rhythm than in networks without rhythm; 2) in networks with internal rhythm while categorizing pieces with the same rhythm as their own; 3) in networks with and without internal rhythm in categorizing computerized rhythm.

HYPOTHESES

In this research we adopted a new continuous technique for implementing musical stimuli in order to verify the following hypotheses:
- Networks with internal rhythm can discriminate between music stimuli characterized by binary rhythm and those characterized by ternary rhythm;
- There is an interaction between the rhythm of the network and the rhythm of the musical piece.

**METHODOLOGY**
- **Stimuli**
The stimuli used for the research were 190 in total, formed by different melodic lines taken in the learning phase from Ars Musica, IV vol.; in the generalization phase, from The Ninth Symphony by L. van Beethoven for binary rhythm and from Nabucco by G. Verdi for ternary rhythm. The choice of real melodies is purely aesthetic. Stimuli were read and translated in numbers by the experimenter directly from musical scores with binary and ternary rhythm. Of these, 100 stimuli (50 with binary rhythm and 50 with ternary rhythm) were used for the training phase and 90, mixed for the rhythmic type, were used in the generalization phase. The input pattern has 64 elements, each one contains a musical note, from C to B, successively divided into two categories: strong-weak musical notes, on the base of the presence or absence of accent, have been represented with odd or even numbers. Each codification must maintain those pieces of information which are remarkable to be reversible, retranslatable, if not totally, at least partly in the original information. We propose the following system:

<table>
<thead>
<tr>
<th>Not accented notes (odd numbers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 3 5 7 9 11 13 15 17 19 21 23</td>
</tr>
<tr>
<td>C C# D D# E F F# G G# A A# B</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Accented notes (even numbers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 4 6 8 10 12 14 16 18 20 22 24</td>
</tr>
</tbody>
</table>

- **The network**
The network used is formed by three (input, hidden and output) unit layers with back-propagation learning algorithm (Rumelhart et al., 1986). The network has 64 input units, 20 hidden units and 1 output unit. We used four networks in order to simulate: 1) the absence of any internal rhythm (RN); 2) a binary rhythm (R2); 3) a ternary rhythm (R3); 4) a pentenary rhythm (R5, composed by 2+3). The difference between k-nary rhythm (k = 2, 3, 5) and rhythm absence is given by the presence of a threshold to the hidden layer unit input, modulated by rhythm changes of (k-nary) type. As regards the threshold choice, a preliminary analysis of output data showed a wide presence of the same values around 0.5. So this value was considered as a useful demarcation to classify the output data in a dual way. The hidden and output units are characterized by a sigmoidal activation function. This function transforms the network input of each unit (the algebraic sum of all excitements and inhibitions, corresponding to a negative value) in the value of its excitement state, that could run from 0 to 1. When the network input is 0 the activation state is 0.5; when the output is greater than 0.5 the input is a binary rhythm, and the input is a ternary rhythm if the output is smaller than 0.5.

- **Simulation test**
A file for each stimulus was built and the 190 total patterns were collected into two principal files. The first, formed by 100 patterns, was used for the training phase. The second file, formed by 90 patterns, was used for the generalization phase. The choice of a 64 elements for input pattern was not casual. The previous attempts were characterized by training phases of back-propagation neural network with a very high cycle numbers. The choice of 20 hidden units has been taken considering the number of the right responses after a certain number of training attempts of the neural network. The learning rate choice was slow and gradual. The error value was modified several times because in the training phase there is a series of input activation patterns and for each one there is also its training input. The network processes this
pattern, several times, sometimes many times. Each time a pattern is activated in input, the activation is diffused to the hidden units and from here to the output unit. The activation pattern on the output unit is compared with the expected teaching vectors and so the error for each output is calculated. At the beginning of the learning the total error is rather high as the weights on the connections were attached in a casual way. The problem for the network is then to reduce this global error, using error on each output unit in order to modify the weights on the connections so on the next trial the total error is lower than the previous one. Learning ends when the total error is lower than that one fixed by the experimenter. In this case, the value of the learning rate used was 0.3. The global error adopted was 0.01. The width of the rhythmic modulation threshold for the network was 0.02.

RESULTS
The results point out a good degree of learning with about 900 training cycles for the four networks. Data in table 2 represent the number of right responses obtained with 90 patterns submitted to the network in the generalization phase.

Table 2: correct responses in the generalization phase:

<table>
<thead>
<tr>
<th>Net type</th>
<th>binary rhythm</th>
<th>ternary rhythm</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>RN</td>
<td>26</td>
<td>26</td>
<td>52</td>
</tr>
<tr>
<td>R2</td>
<td>34</td>
<td>25</td>
<td>59</td>
</tr>
<tr>
<td>R3</td>
<td>24</td>
<td>33</td>
<td>55</td>
</tr>
<tr>
<td>R5</td>
<td>25</td>
<td>24</td>
<td>49</td>
</tr>
</tbody>
</table>

chi square= 2.81  \( p= 0.4 \)

As we can see from these results networks with internal binary or ternary rhythm show a better behaviour than the network with pentenary rhythm and the one without rhythm. However, both R2 and R3 networks discriminate stimuli with a rhythm in accordance with their own better than stimuli with the opposite rhythm. The pentenary rhythm network does not facilitate the discrimination neither for the binary rhythm and nor for the ternary rhythm. Finally, the network without rhythm does not show any remarkable difference in the correct responses between binary and ternary passages.

Since the chi square test is not statistically significant, we enlarged the network sample. We formed three nets groups: 15 networks without rhythm, 15 with ternary rhythm and 15 binary rhythm varying the input weights casually. At first, it was verified if the inside behaviour of the three groups of networks was coherent (see table 3).

Table 3: test of the behaviour coherence inside each network group.

<table>
<thead>
<tr>
<th>Nets type</th>
<th>N°</th>
<th>chi square</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>RN</td>
<td>15</td>
<td>1.94</td>
<td>.99</td>
</tr>
<tr>
<td>R2</td>
<td>15</td>
<td>1.76</td>
<td>1</td>
</tr>
<tr>
<td>R3</td>
<td>15</td>
<td>1.61</td>
<td>1</td>
</tr>
</tbody>
</table>

As networks within each group do not behave in a significantly different way, the results of each group were considered cumulatively. In this case, a very significant chi square test was obtained in the total sample formed by 45 networks (chi square test= 23.831; \( p=.0001 \)). The variability derives from an initial, specific and casual weight assignment for each network because learning consists in the modification of these weights, its course and final results vary from network to network. By repeating the experiment with different initial weight assignment results are similar but not identical. Varying frequency and order of the patterns, learning takes place anyhow. In order, to evaluate the correlation between network performance and pattern length, the number of the network input units has been reduced and correlated to the number of correct recognition. The number of 64 input units has been progressively reduced maintaining variables unchanged. The number of cycles necessary for the training phase of the back-propagation neural network is nearly constant up to 54 input units. Further reduction of the
number of input units produced an increasing of the training phase according to the following table.

Table 4

| Cycles with 54 input units | 971 | 953 | 1008
|---------------------------|-----|-----|------
| Starting from 49 input units the number of cycles has remarkably increased reaching values greater than 5,000 - 6,000 for each training phase. From 44 input units onward, the number of training cycles is even greater and a very slow reduction of error is evident (see table 4).

Table 5: general results

<table>
<thead>
<tr>
<th>With 54 input units:</th>
<th>RN 27</th>
<th>RN 26</th>
<th>RN 24</th>
<th>RN 22</th>
<th>RN 24</th>
<th>RN 26</th>
<th>RN 24</th>
<th>RN 22</th>
<th>RN 24</th>
</tr>
</thead>
<tbody>
<tr>
<td>bin</td>
<td>tern</td>
<td>tern</td>
<td>tern</td>
<td>tern</td>
<td>tern</td>
<td>tern</td>
<td>tern</td>
<td>tern</td>
<td>tern</td>
</tr>
<tr>
<td>tot</td>
<td>53</td>
<td>53</td>
<td>53</td>
<td>53</td>
<td>53</td>
<td>53</td>
<td>53</td>
<td>53</td>
<td>53</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>With 53 input units:</th>
<th>RN 26</th>
<th>RN 25</th>
<th>RN 24</th>
<th>RN 22</th>
<th>RN 24</th>
<th>RN 26</th>
<th>RN 25</th>
<th>RN 24</th>
<th>RN 22</th>
</tr>
</thead>
<tbody>
<tr>
<td>bin</td>
<td>tern</td>
<td>tern</td>
<td>tern</td>
<td>tern</td>
<td>tern</td>
<td>tern</td>
<td>tern</td>
<td>tern</td>
<td>tern</td>
</tr>
<tr>
<td>tot</td>
<td>51</td>
<td>51</td>
<td>51</td>
<td>51</td>
<td>51</td>
<td>51</td>
<td>51</td>
<td>51</td>
<td>51</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>With 52 input units:</th>
<th>RN 24</th>
<th>RN 24</th>
<th>RN 24</th>
<th>RN 24</th>
</tr>
</thead>
<tbody>
<tr>
<td>bin</td>
<td>tern</td>
<td>tern</td>
<td>tern</td>
<td>tern</td>
</tr>
<tr>
<td>tot</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>49</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>With 51 input units:</th>
<th>RN 25</th>
<th>RN 24</th>
<th>RN 23</th>
<th>RN 23</th>
</tr>
</thead>
<tbody>
<tr>
<td>bin</td>
<td>tern</td>
<td>tern</td>
<td>tern</td>
<td>tern</td>
</tr>
<tr>
<td>tot</td>
<td>51</td>
<td>51</td>
<td>51</td>
<td>51</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>With 50 input units:</th>
<th>RN 24</th>
<th>RN 23</th>
<th>RN 23</th>
<th>RN 23</th>
</tr>
</thead>
<tbody>
<tr>
<td>bin</td>
<td>tern</td>
<td>tern</td>
<td>tern</td>
<td>tern</td>
</tr>
<tr>
<td>tot</td>
<td>48</td>
<td>48</td>
<td>48</td>
<td>48</td>
</tr>
</tbody>
</table>

**DISCUSSION**

The sample learning experiment confirms the superiority of networks with internal binary or ternary rhythm with respect to the network without rhythm and the pentenary one in processing rhythmic inputs. A better discriminative ability is shown by networks with inner binary or ternary rhythm towards the compatible input pattern, while the other two classify indifferently binary or ternary inputs. However, results given by networks are never completely wrong or right, instead there is an approximation in one sense or in another: that is, after learning networks never give an activation value of 1 or 0 but values, for example, of 0.93 or 0.2. The replica with a weight change in network groups, each one formed by 15 units, showed an interesting feature of neural networks, which can give the same result with a variety of weight inner values. It is not important a certain weight, but all weights on all connections. It would be of great interest to analyse the behaviour in the light of a larger investigation about other possibilities in the network field.
REFERENCES
Beethoven L. Nona Sinfonia. Peters, Germany.
Verdi G. Nabucco. Ricordi, Italia
INTRODUCTION

Previous researches show that:
- music is a language with its own syntax and semantics (Sloboda 1988);
- the most fruitful learning of musical language is that occurring along with the ability development (Olivetti Belardinelli 1997);
- deaf children have normal intelligence (AlES 1975; 1994) and show a normal cognitive development when receiving a specific education based on visual and physical stimuli;
- deaf children have musical cognitive abilities on the same level of hearing ones (Olivetti Belardinelli 1990).

Starting from these considerations we performed a longitudinal research project in order to verify the ability of deaf children to develop musical behavior, that is to perform and to produce music as consequence of a fit and qualified musical training.

METHOD

The research was performed during 3 schoolyears from 1993 to 1996 in a special school for deaf children in central Italy.

Test-retest sessions were performed at the beginning and at the end of each schoolyear; during each schoolyear children received regular musical training.

The total number of trained children varied in each year according the following scheme:

<table>
<thead>
<tr>
<th>Schoolyear</th>
<th>N.of Ss</th>
<th>nursery school</th>
<th>primary school</th>
<th>middle school</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993/4</td>
<td>11M+4F</td>
<td>11</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>1994/5</td>
<td>16M+4F</td>
<td>5</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>1995/6</td>
<td>12M+4F</td>
<td>3</td>
<td>8</td>
<td>5</td>
</tr>
</tbody>
</table>

Only 7 boys and 4 girls attended the music lessons during all 3 years and for them it was possible to perform a follow-up.

Test-retest was performed by means of the Musical Behavior Scale of Olav Skille (MUBS). The MUBS is based on Skille's concept of musicality as a primary human function in a framework of general functions, which are necessary to be able to give musicality a musical expression. Therefore the MUBS concentrate on musical behavior that, according to the author's view, is closely dependant on intelligence, personality, motor and social functions. The necessity of assessing the subject's musical behavior rules out group testing and the testing session develops, along with the communicative interactions between experimenter and subject according to the following modalities of communication: 1) Rhythmical; 2) Dynamic; 3) Melodic; 4) Auditive; 5) Improvisatoric; 6) Vocal; 7) Movement. In the adaptation we made for use with deaf children the 4th and 6th subtests were omitted. Necessary instruments for MUBS administrations are two drums, a xilophone, a tape recorder.

During the schoolyear deaf children received regular musical lessons in fit music classrooms, while staying on wooden footboards. Cognitive, affective and psychomotor areas were particularly trained. We devised special training exercises according to the main scope of the lesson: for instance, different colours were used to highlight differences in pitch, intensity or duration; pitch and sound recognition were trained by means of teacher's playing 3 different plates (c3, f, c4) while children, sitting on the wooden footboard, maintained their hands on the board; after this special training blindfolded children could recognize the different sounds of the different plates.
RESULTS
At the end of the observation period the children know the musical syntax and semantic and are able to play together in time and in tune and to compose little pieces for drums in the primary school, and for flute in the middle school. All of them are more attentive to the sound environment.

MUBS scores improve at the end of each year as one can see in table 1, where it is evident that rhythmical, melodic and movement scores showed the greater improvement.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rhythmical</td>
<td>19,333</td>
<td>23</td>
<td>18,75</td>
<td>21,8</td>
</tr>
<tr>
<td>Dynamic</td>
<td>5,533</td>
<td>5,267</td>
<td>4,65</td>
<td>5,55</td>
</tr>
<tr>
<td>Melodic</td>
<td>14,067</td>
<td>16,6</td>
<td>11,25</td>
<td>14,4</td>
</tr>
<tr>
<td>Creative</td>
<td>2,933</td>
<td>2,733</td>
<td>2,4</td>
<td>2,45</td>
</tr>
<tr>
<td>Body Movement</td>
<td>4,867</td>
<td>5,067</td>
<td>3,5</td>
<td>4,45</td>
</tr>
<tr>
<td>Total Musical</td>
<td>47,8</td>
<td>53,467</td>
<td>41,25</td>
<td>49,5</td>
</tr>
</tbody>
</table>

Afterwards we undertook the comparison between the scores obtained by the children who get the music courses all 3 years along, respectively at the beginning of the training and after the complete 3 years training. Here we have only 11 subjects: 6 of them at the end of the 3 years were still in the primary school, while 5 (4 boys and 1 girl) entered the middle school at the beginning of the second year of training. Table 2 shows an improvement for all quotients excepts for the dynamic, creative and motor ones.

<table>
<thead>
<tr>
<th>Quotient</th>
<th>1993</th>
<th>1996</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rhythmical</td>
<td>15,545</td>
<td>25,182</td>
</tr>
<tr>
<td>Dynamic</td>
<td>5,364</td>
<td>5,818</td>
</tr>
<tr>
<td>Melodic</td>
<td>12,273</td>
<td>17,091</td>
</tr>
<tr>
<td>Creative</td>
<td>2,909</td>
<td>2,909</td>
</tr>
<tr>
<td>Body Movement</td>
<td>4,727</td>
<td>4,727</td>
</tr>
<tr>
<td>Total Musical</td>
<td>42,182</td>
<td>56,545</td>
</tr>
</tbody>
</table>

Finally we performed a cross analysis of the course of the considered musical abilities in the 3 school levels, during the 3 years training (table 3: here again all subjects are considered).

We found a general improvement of the Total Musical Quotient, and in details a high improvement of the rhythmical quotient in the primary school and of the melodical quotient in the middle school, while the dynamic, creative and motor quotients don’t exhibit appreciable changes.
## RHYTHMICAL QUOTIENT

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>nursery</td>
<td>-</td>
<td>15,54</td>
<td>6,4</td>
<td>8,2</td>
</tr>
<tr>
<td>primary</td>
<td>29,75</td>
<td>20,09</td>
<td>11</td>
<td>18</td>
</tr>
<tr>
<td>middle</td>
<td></td>
<td>31</td>
<td>30,77</td>
<td>31,88</td>
</tr>
<tr>
<td>class differences</td>
<td>p=,0066</td>
<td>p=,0001</td>
<td>p=,0001</td>
<td></td>
</tr>
</tbody>
</table>

## DYNAMIC QUOTIENT

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>nursery</td>
<td>-</td>
<td>5,36</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>primary</td>
<td>6</td>
<td>5</td>
<td>4,3</td>
<td>6</td>
</tr>
<tr>
<td>middle</td>
<td></td>
<td>6</td>
<td>5,7</td>
<td>5,8</td>
</tr>
<tr>
<td>class differences</td>
<td>p=,043</td>
<td>p=,0066</td>
<td>p=,0027</td>
<td></td>
</tr>
</tbody>
</table>

## MELODIC QUOTIENT

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>nursery</td>
<td>-</td>
<td>2,90</td>
<td>2</td>
<td>2,4</td>
</tr>
<tr>
<td>primary</td>
<td>3</td>
<td>2,72</td>
<td>2,16</td>
<td>2</td>
</tr>
<tr>
<td>middle</td>
<td></td>
<td>2,75</td>
<td>2,77</td>
<td>2,77</td>
</tr>
<tr>
<td>class differences</td>
<td>p=,8818</td>
<td>p=,0017</td>
<td>p=,411</td>
<td></td>
</tr>
</tbody>
</table>

## CREATIVE QUOTIENT

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>nursery</td>
<td>-</td>
<td>2,90</td>
<td>2</td>
<td>2,4</td>
</tr>
<tr>
<td>primary</td>
<td>3</td>
<td>2,72</td>
<td>2,16</td>
<td>2</td>
</tr>
<tr>
<td>middle</td>
<td></td>
<td>2,75</td>
<td>2,77</td>
<td>2,77</td>
</tr>
<tr>
<td>class differences</td>
<td>p=,8818</td>
<td>p=,0017</td>
<td>p=,411</td>
<td></td>
</tr>
</tbody>
</table>

## BODY MOVEMENT

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>nursery</td>
<td>-</td>
<td>5,36</td>
<td>4,2</td>
<td>4</td>
</tr>
<tr>
<td>primary</td>
<td>6</td>
<td>5,27</td>
<td>3,66</td>
<td>4,16</td>
</tr>
<tr>
<td>middle</td>
<td></td>
<td>4,5</td>
<td>3</td>
<td>4,66</td>
</tr>
<tr>
<td>class differences</td>
<td>p=,258</td>
<td>p=,7859</td>
<td>p=,3495</td>
<td></td>
</tr>
</tbody>
</table>

## TOTAL MUSICAL QUOTIENT

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>nursery</td>
<td>-</td>
<td>42,18</td>
<td>19</td>
<td>24</td>
</tr>
<tr>
<td>primary</td>
<td>63,25</td>
<td>48,7</td>
<td>29,16</td>
<td>42,66</td>
</tr>
<tr>
<td>middle</td>
<td></td>
<td>66,5</td>
<td>61,66</td>
<td>68,22</td>
</tr>
<tr>
<td>class differences</td>
<td>p=,0148</td>
<td>p=,0001</td>
<td>p=,0001</td>
<td></td>
</tr>
</tbody>
</table>
DISCUSSION

Results show that a fit training improves deaf children’s musical ability. Comparing the data of MUBS collected at the beginning of 1993 with the data collected at the end of 1996 we find a significant increase in the rhythmical and melodic quotients and in the Total Musical Quotient.

This result is more evident in the longitudinal comparison 1993-1996.

As regard age differences we found that musical training is more effective during the nursery and the primary school, confirming by this way that the age of the development of abstract intelligence is the fittest in order to learn the musical language.

As abstract intelligence is consolidated only the ability in managing musical language as assessed by the melodic quotient undertake a substantial improvement: we believe that this result derives from the strengthening of attention and memory as necessary conditions to repeat difficult melodical items.

As a conclusion we will understriken that: 1) deaf children, if opportunely trained, develop musical cognition in the same way and at the same ages than normally hearings subjects and 2) the importance of music education for deaf children is that of fostering a complete cognitive development, giving to this subjects another communication possibility.

REFERENCES

KARMA KAI (1994). Auditory and visual temporal structuring: how important is sound to musical thinking? Special Issue: Assessment in music. In Psychology of music; Vol 22(1) 20-30
SKILLE, O. (**). The Musical Behaviour Scale (MUBS): Theoretical Background and Description. Ciclostilato.
Retrieving Long Term Memory traces in contemporary music listening: 
a composer's view

Fabio Cifariello Ciardi
Conservatorio "L.Perosi" - Campobasso
via Pietro Giannone, 28 - 00195 - Roma ITALY
e-mail: f.cifariellociardi@agora.stm.it

1. Introduction
Human memory represents a central area of psychological research and, in general, a fundamental issue for understanding human communication processes. Composers might be interested in evaluating listeners memory processes in order to optimize communication of their aesthetic and poetic ideas. However, composers' consideration of memory is often ambiguous, as while they understand that the possibility to grasp structural invariance of a piece of music depends on short-term memory (STM) and long-term memory (LTM) interactions, nevertheless they seldom examine the role of LTM in musical experiences closely. LTM refers to knowledge which is already stored while composers ideally tend to explore original, unknown territories.

Many reasons might justify composers' attitude towards Tulving's distinction of LTM in episodic and semantic memory. Episodic memory is scarcely significant in composition because of its mainly autobiographical nature (i.e., episodic memory is a personal, unique wisdom which often looses its consistency when applied to large groups of people); on the other hand, remarks about semantic memory are often avoided in order to refrain from facing the puzzling issue of musical semantics.

Today, however, reconsidering the influence of LTM on music listening might be useful for aesthetic, ethical and even political reasons (Cifariello Ciardi, 1994), but even simply for a better evaluation of the cognitive "weight" of music materials. Furthermore, New Media urgently call for such an evaluation: a better understanding of storing and retrieving processes is clearly helpful in the organization of the huge sonic universe now available by means of digital technology.

2. Denotation, connotation, commonness of sonic events
Composers are more interested in the unstable zone between expected and unexpected sonic events, rather than in the listener's recognition of specific musical items. What composers might wish to investigate is the possibility to influence, by means of appropriate compositional techniques, the way listeners retrieve LTM traces. That is, composers are interested in controlling the semantic quality of sonic events in order to improve the "depth" of their acoustic space (i.e., a depth defined by a figure-ground perspective). This kind of control might be helpful in shaping subject's attention, strategies and expectations while listening to music genders based on non-traditional materials and syntax.

In order to assign semantic qualities to an acoustic stimulus, sensorial processing must activate structures of semantic components stored in LTM. As LTM structures and elements may be represented by semantic networks, activation of previously learned information may be described as a process of semantic features correlation. Within this framework, the "meaning" of sonic events is the result of structures of semantic components which may emerge from such a correlation process. From a semiological point of view, by means of this process, the subject may interpret a stimulus as a sign (i.e., a mental construct provided with a specific meaning). The mental representation of the sign will primarily lead the subject to denotate the stimulus, giving by this way an immediate meaning, and subsequently connotate the same stimulus giving a secondary meaning that arise from the first.

Denotation might refer to the result of the very first correlation of the sign with previously learned knowledge of the subject, while connotation might be considered as the result of subsequent correlations. Thus far, denotation and connotation have been used here in a quite broad sense: they indicate the result of any semantic features correlation with sensory data, rather than specific goals of semantic encoding (e.g., recognition of the cause that produces the stimulus,
identification of the function of the stimulus within a certain context, recall of an appropriate noun).

Strictly speaking connotation of a sonic event is quite a subjective attribute; nevertheless, if a stimulus would activates - among others - analogous semantic features within a group of subjects, we might hypothesize that a commonness in subjective connotations was elicited. The eventual emergence of connotation commonness depends on the number of subjects that associate the stimulus to a common set of semantic components. Moreover, connotation commonness might be related to the number of interconnected semantic features: the more interconnected semantic features activated by the stimulus are, the more evident the strength of connotation will be and the stronger commonalities will emerge. For example, single tones of well-known musical instruments might give rise to simple common connotations (e.g., the common connotation of a piano tone might refers to the musical instrument that produced the sound), while more complex shared connotations are difficult to be considered.

On the other hand, a stronger convergence towards more interconnected semantic features might be expected in case of different types of sounds, although immediate correlations might still be difficult to foresee.

Considering a well-known musical theme, a scream, a car beep, the sounds of wind or rain, a shared connotation might be predicted. Such connotation might be based on a relatively small number of common semantic components highly interconnected in some analogous way within the memory network of each member of the group. Commonalities among subjective connotations might refer to various type of cognitive activity: abstract or procedural knowledge, emotional or motorial response. However, what remain clear is that different cognitive (and biological) structures may lead to similar cognitive responses, thus implicating somehow analogous semantic encoding (e.g., different subjects might regard a sonic event as a shared concept more than subjective percept).

Our claim is that strong common connotation of any sonic event implies encoding modalities and semantic effects closely related to those that arise in long-term verbal memory.

Connotated sonic events have hardly been studied as a whole. However, since well-known melodies may be considered as a commonly connotated sonic event, results from literature on well-known melodies memory representation (e.g., Dowling & Harwood, 1986; Halpern, 1984) may be used to evaluate the aforementioned assumption. Among others, evidence for a parallelism between well-known melodies and word-list encoding modality comes from experiments by Gardiner et al. (1990). Their results show that the stable relation between recognition and recall (Tulving-Wiseman law) is not restricted to verbal material, but is extended to familiar melodic excerpts.

Considering commonness connotation as an index of a class of sonic events, two distinctive features ought to be emphasized. First of all, such a class of connotated sonic event will include a wide range of sounds produced by very different media: sounds from a natural environment as well as sound produced by human being or objects for various reasons. Secondly, connotation can be described on a relative scale only. That is, connotation refers to commonalities that emerge within a group of subjects (i.e., a reference group) that agree on specific semantic qualities of an event.

Consequently, common connotation of a sonic event might be defined on a statistic basis by taking into account its stability in time and space domain: common connotation of a sonic event may be strong for a certain period of time and weaker after that time. For instance, the sound of horse-drawn coach was probably much more connotated before Industrial revolution than in the present-day word. Nevertheless, despite the uncertain edge of a class of commonly connotated sonic events, the definition of the issue seems to be plausible throughout analysis of sonic landscape invariances (cf. Murray Schafer, 1977). A higher rate of constancy and invariance within a certain time and a certain space will probably lead to a stronger and more common connotation. In this sense, contemporary world represents an exceptional ground for sonic connotation inquiries: common auditory experiences in the present-day multimedia society are frequent; this means that, today more than in the past, we might expect to encode sonic experiences in a more tangible common memory network.
3. Cognitive parameters in sonic connotation

Composers might gain several advantages from investigation on connotated sonic events cognitive processing. If semantics plays a crucial role in listening to a connotated sonic event it means that composers could set up an appropriate compositional processes in order to take advantage of effects related to semantic encoding of sonic materials. As an example, experimental evidence on semantic encoding of verbal items points out that our cognitive system produces faster responses for semantically encoded information (e.g., listener cognitive processing of recognizable inputs is faster than cognitive processing of non-recognizable inputs) (cf. Chang, 1986 for a review). Hence, if known aspects of a sonic stimulus are processed faster than unknown ones, than composers may control processing time in music perception through connotation control. That is, composers might use connotation in such a way as to stimulate and convey listener's attention towards specific aspects of the musical piece. In order to do this, a central issue regards the physical correlates of a specific semantic quality. The question than arise as to what extent parameters could be modified leaving connotation relatively intact.

The modification of any physical parameter of a connotated sonic event may result in a variation between two boundaries. On one side, connotation strength may remain almost unaffected by parameter modification; that is, modified sonic events will activate semantic features similar to those activated by the original connotated event. On the other side more dramatic modification of the same physical dimension might irrevocably corrupt sonic event connotation; this means that closeness among semantic features of both modified and original sonic events will not be detected. Therefore, we may argue that a physical parameter might erase sonic event connotation whenever parameter modification exceeds a certain threshold. Connotation control in composition depends on the localization of such connotation thresholds. Conditions which determine connotation thresholds may refer to subject's timeless knowledge as well as to time-dependent factors which influence attention or emotional state. Since consistent commonalities among subjective variables are unpredictable, composers tend to ignore the issue. However, connotation thresholds may also depend on subject-independent factors. Acoustical attributes and context distinctive features may contribute to define connotation thresholds. Within this framework composers may wish to control those parameters that seems to be critical in order to common connotation detection; that is, they may want to control parameters which seem to be perceptually salient for any member of the reference group. Therefore, we may pose the following question: can we determine a common "connotative weight" of a sonic event acoustic parameter?

Stable relations between connotation and its physical correlates are difficult to establish. Connotation of a percussion tone depends on amplitude envelope rather than on event length; on the contrary, considering wind sound connotation, we might tend to assign a higher "weight" to the event length and a lower "weight" to the amplitude envelope. Moreover, a straightforward application of theories about music cognition is not always useful to answer the question. For instance, "connotative weight" of pitch transposition seems to change over different sonic event categories. Empirical findings show that relative pitch chroma plays the most important role in memory for familiar melodies rather than absolute pitch (W. J. Dowling & D. L. Harwood, 1986 for a review). These results suggest a low "connotative weight" for pitch transposition. However, for other sonic categories (e.g., animal sounds) we may assume that absolute pitch information is more accurately stored in LTM. If this is the case, absolute pitch transposition would easily corrupt sonic event connotation.

Generally speaking, "connotative weight" of acoustic parameters may be assigned on a cognitive-efficiency basis: critical parameters in sonic event connotation might be those which could lead the subject towards an efficient answer to the expectations usually correlated with that sonic event. This assumption may explain the reason why less significant parameters in musical experiences, such as amplitude and sound localization, become critical in natural environments.

Apart from subjective variables, connotation arousal is highly correlated with subject's exposure time to the stimulus. Under certain conditions connotation almost immediately emerges, while in other situations a longer time-span seems to be necessary to guarantee a reliable semantic qualification. Consequently we may argue that, since connotation stems from semantic correlation activity, connotation assignment may be considered as a time-dependent process: the longer the
exposure to a sonic event, the stronger the connotation will be. This suggests that by means of an exposure time control, composers could control the chance that a connotation assignment will take place. Moreover, whether sufficiently long exposure to a sonic event may give rise to shared connotations, shorter exposure to the same event will probably activate episodic and semantic encoding which leads to much more subjective responses.

As previously mentioned, connotation assignment is influenced by external conditions. Context-dependent factors may influence connotation thresholds through a sort of semantic masking. Under certain conditions it is possible that interferences in semantic encoding will take place when a connotated sonic event is either preceded, or followed, or superposed to an other sonic event. In this case connotation strength might depend on both subject-dependent and subject-independent factors. Presumably, the more complex and robust the semantic network that represents the connotated event is, the more interference-resistant the connotation will be. Yet, the possibility to preserve connotation will be related to the subject's attention and expectations: as classical Cocktail Party phenomenon suggests, we may expect that subjects can almost “shadow” or ignore one sonic event while processing the other one. On the other hand, selective attention and semantic masking are probably related to subject-independent factors as well. For instance, structural coherence between sonic events seems to play a central role in connotation detection. The more acoustic analogies exists between two differently connotated events, the more difficult the single connotation detection will be. From a compositional point of view, semantic masking is an other controllable process that might be used to shape sonic materials connotation.

4. Shaping sonic connotations in musical composition
In one of author's pieces, *Altre Tracce* for Bb clarinet (Cifariello Ciardi, 1991-2), exposure time and semantic masking have been considered with the purpose of affecting listeners' semantic encoding. Two groups of well-known melodic excerpts have been used as connotated sonic events. The first group includes, among others, the opening theme from W. A. Mozart's K.550 symphony, the main theme of G. Rossini's *Barbieri di Siviglia* and the second theme from Rossini's *La Gazza Ladra* overture. The common feature among all excerpts of the former group is a pattern of two or more repeated notes embellished by a descending and an ascending minor second. The second melodic group includes the *Seguidilla* from G. Bizet's *Carmen*, a prominent secondary theme from Debussy's clarinet and piano *Premiere Rapsodie* and the jazz standard *The Lady's a Tramp*. The common feature among all excerpts of the latter group is an embellished segment of a descending chromatic scale. The piece proposes continuous oscillations among connotated melodic fragments and more “neutral” gestures based on a 12-tone row derived by merging the two groups of excerpts.

In the first section of the piece two sonic streams are alternatively presented. The first stream tends to break over the *Barbieri di Siviglia* theme four time, while the second stream points at Bizet's *Seguidilla*. First and second (Example 1) appearance of *Barbieri di Siviglia* theme is normally too short in order to permit recognition or connotation detection. In the successive exposition the same theme is still obscure because of the exclusion of the initial three-notes segment (Example 2). However, in both cases the sonic stream is always broken over in the same pitch region. The goal of the polarization is to generate listeners' expectation associated to a specific register.

During the third exposition listeners often recognize the theme. This means that the excerpt connotation has emerged from nearby sonic materials (Example 3). Here, connotation detection may depend on three factors. Firstly, the length of the third exposition is such that common correlations emerges. Secondly, previous Rossini presentations contribute to the subject's pre-activation. Thirdly, semantic masking is avoided by means of contrasting structures of standing by events.

An important secondary effect of connotation assignment in music listening is a sort of semantic “dissonance” between a connotated events and other neighboring ones. That is, semantic encoding of the two events may activate concept nodes which are semantically far one from the other (e.g., interconnection between connotate and “neutral” events memory networks may be difficult to establish). Consequently, a strong loss of global coherence may be detected by the
subject. Hence, where efforts have been made to facilitate connotation assignment, other compositional techniques have been used to smooth passages between connotated and more "neutral" events. Firstly, hidden anticipation of Rossini's theme's salient features (Example 1 and 2) has been intended as a "preparation" to the third recognizable exposition of the excerpt. Secondly, since pitches of the connotated sonic materials are embedded in the "neutral" 12-tones row, collage-effects might be attenuate if underlying structural links are detected. Once connotation has been confirmed, it has been faded out by means of a progressive semantic masking.

Finally, according to author's empirical results connotation may play a role in stream segregation. The typical loss of a sense of temporal order across two different streams seems to increase when connotation of the two streams is different. This assumption has been tested in an other section of the piece (Example 4). Listener responses suggest that connotation may be used as an extra-cue in interleaved melodies identification. This result is coherent with Hartmann and Johnson conclusions (Hartmann and Johnson, 1991): although peripheral channeling is the dominant characteristic in stream segregation, other central processes - as semantic encoding of connotated sonic events - might arise.

4. Conclusion
This paper is an attempt to point out how and why composers and music theorists should thoroughly examine the encoding modalities of well-known sonic events. The result of the author's empirical studies leads to the following conclusions.
1) Connotation can be usefully considered to index semantic encoding of a relatively large class of sonic events.
2) Commonness among subjective connotations may be predicted though a statistical sonic landscape analysis, and may lead to the definition of a class of commonly connotated sonic events.
3) For each acoustic parameter of a connotated event, a threshold may be assumed in order to define to what extent parameter modification will not corrupt connotation assignment.
4) Digital Sound processing as well as appropriate compositional techniques may supply creative control on exposure time, masking effects and other factors related with connotation detection.

We are aware that many of the aforementioned assumptions are hypothetical and speculative ones and they need to be corroborated by a great amount of theoretical evaluation and empirical research. Nevertheless we hope that they might renew psychologists interest in the study of contemporary and electronic music sonic materials.

5. Acknowledgments
The author would like to thank Prof. Olivetti Berardinelli for her valuable criticism to the manuscript.

6. References
IS MUSIC UNAVOIDABLE?

P. Camiz
ECONA and Department of Physics-Università di Roma "La Sapienza"

From a strictly technical point of view, music can be considered as an organized walk on a discrete set of points in the otherwise continuous space of frequencies: there are few exceptions (glissando, vibrato, portamento) which allow the invasion of this continuous domain, but the major part of everybody's musical experience is developed by means musical pitches. Why? I can find at least two good reasons for this choice: although asemantic, music, as a language, needs to be robust against distortions occurring during performance, transmission and reception. This condition is easily satisfied if the elements of the language considered are "distant" from each other in their space. The second reason is related to the particular structure of many natural sounds, among which the human voice: the occurrence of harmonic overtones, whose frequencies are integer multiples of a fundamental pitch. This circumstance is responsible for the privilege given to intervals corresponding to the ratio of small integers. Moreover the fact that music is based on ratios, rather than differences, of frequencies, makes it invariant with respect to relative motion between source and receiver. The Doppler effect is another case of invasion of the continuous by the rigid and discrete structure of musical pitches.

Musical theories are mostly grounded on numerological arguments, involving small integers, or physical ones, which consider harmonic overtones, but they devote scarce attention to the auditory apparatus. It is necessary to look more carefully at the role played by each of the three sections in which this apparatus can be logically divided: the intermediate ear, the cochlea and the auditory cortex. Let's start with the cochlea, which can be considered as an active analogical-digital converter, where the mechanical stimulus is first transformed into a global excitation of the fluid and of the basilar membrane, and then into the localized electrical excitation of the acoustic nerve. It is very appealing to look at the inner ear as to a discrete Fourier transformer, although it probably differs in many ways from a mathematical one; effects of deviations from correct behaviour can be taken into account when discussing the auditory cortex, which shows experimentally the so called tonotopic organization: neurons tuned to a given frequency are roughly located in a position which depends almost linearly on the frequency logarithm.

Surprisingly enough, the auditory cortex is much less known than the visual cortex, but some general properties can be accepted, namely that synaptic connections can develop under the influence of auditory stimuli. Their structure in the frequency domain can be easily described, e.g. in terms of a correlation matrix, where all the individual auditory experience is condensed. Such a matrix exhibits a gross structure, produced by sounds with a continuous spectrum, and superimposed a fine structure, corresponding to sounds with a discrete spectrum. It is this second part which can act selectively on the synaptic development: pairs of neurons which are excited simultaneously, because their frequencies are common overtones of a fundamental pitch, are more probably connected by an excitatory synapse rather than by an inhibitory one. There are many models describing the process by which the auditory cortex can reach its definite organization: one, following Oja's ideas, uses eigenvectors of the correlation matrix as a representation of synaptic connections between a first layer of neurons and a second one, whose units
are related to the principal components of the ensemble of input vectors (P. Camiz SMAC93); another one looks at the topological organization of a self-organizing neural network of Kohonen type produced by the correlation matrix (Camiz, Beato ICA95). In both cases consonant intervals are privileged, suggesting that the development of music is due to the "harmonic structure" of the correlation matrix, and to the learning mechanism of the auditory cortex.

I want to discuss now the role which can be played by the medium ear, namely the mechanical system composed by the tympanic membrane, the ossicular chain and the oval window: due to the enormous range of the excitation amplitudes (about 5 orders of magnitude) even a small anharmonic term in the system should be revealed, at least for stimuli of high intensity, through the non linearity of its response to a simple stimulus. In order to study the consequences of this term it suffices to describe the whole system as a one-dimensional anharmonic, damped, driven oscillator, whose equation of motion is
\[ m \cdot a + k \cdot f(x) + s \cdot v = F(t), \]
where \( x, v, a \), are respectively the displacement, the velocity and the acceleration, \( m \) the mass, \( k \) the elastic constant, \( s \) the damping term, \( f \) any odd monotonous function of \( x \), and \( F(t) \) the driving term. If \( f(x) = x \) we have the harmonic oscillator, if \( f(|x|) \) is greater (smaller) than \( |x| \) the rigidity increases (decreases) with the amplitude, and we can expect that the free oscillation's frequency be greater (smaller) than the harmonic one \( \omega_0 = \sqrt{k/m} \). This can be shown analytically in some cases, like the simple pendulum, and in any case by numerical integration of the equation of motion.

I want to show some of the results I have obtained with \( f(x) = \sinh(x) \), in different cases, namely: free oscillations \( (F(t) = 0) \), simply or douply periodic driving force \( (F(t) = \sin(\omega_1 \cdot t), F(t) = \sin(\omega_1 \cdot t) + \sin(\omega_2 \cdot t)) \). Free oscillations show clearly (fig1) the increase of frequency with the amplitude, accompanied by the appearance of odd overtones. When a purely sinusoidal driving term is introduced and the damping is conveniently chosen, the response of the system soon becomes purely sinusoidal for small amplitudes, and is richer in higher overtones of the driving frequency when its amplitude rises (fig. 2). This is already an important result, because it indicates that even in absence of overtones in the driving term, the internal correlation matrix may exhibit its fine structure discussed in the preceding paragraphs, provided that the stimuli are strong enough. When there are two driving terms another phenomenon occurs, which has also musical relevance. Contrary to the linear case, in which the superposition principle holds, two (or more) driving terms give rise to a response which is not the sum of the single ones (enriched by the overtones); the result is more complex, and suggests a new criterion for the evaluation of consonance.

The sum of two (or more) simply periodic functions has a period given by the m.c.m. of the individual periods, which is finite if all the periods are in rational ratios, and is shorter the smaller the integers forming the ratios. Correspondingly there is a "virtual" fundamental frequency, whose amplitude may be zero in the driving term, and remains zero in the case of linear response; when the response is non linear, and the driving terms are large, the virtual fundamental becomes excited, together with all its overtones, among which the driving frequencies are obviously included. In fig3 some examples are shown of two driving sinusoidal terms, together with the responses of the non linear system when the ratio is given by small or large integers. The Fourier amplitudes of the virtual fundamental and of its overtones become important for higher excitation, while the density of excited frequencies becomes larger for longer overall
period of the driving force. A quantity related to the spectral density of the response may be chosen as a measure of the consonance, if one remembers that such a response is the exciting term of the cochlea. This can be synthesized by the trivial sentence: "It is much easier (for the auditory cortex) to deal with simple things rather than with complicated ones".

If this hypothesis is correct, it turns out that our auditory system, when working in a non linear regime, adds something to the external stimulus, and we hear "acoustical ghosts" like Tartini's third sound; moreover, if the ratios of the driving periods are irrational, the ghosts invade the whole spectrum and the dissonance is total. In most cases we are not aware of such a distortion, simply because the stimulus is already dressed with its harmonic overtones, and the only effect of the non linear response is to modify more or less their amplitudes and phases; the resulting perception is probably a sound whose timbre changes with its intensity.

During the evolution of the musical language we became more and more tolerant to dissonances, and today's composers, sometimes with the help of electronics, but also with normal musical instruments, often propose music based upon an infinity of elements, rather than the traditional 12 notes. The task of the listener becomes more and more difficult, if we accept the idea that our auditory system has been trained in an acoustical environment dominated by periodic stimuli. It is impossible, for a finite mind, to identify an element belonging to an infinite ensemble, and this is true also for the composer! It seems reasonable that a finite mind, exposed to this kind of music, makes a reduction to finiteness, defining a large but finite number of classes with the help of its personal experience: each class becomes a symbol of the listener's language, whereas elements belonging to the same class constitute the ensemble in which the freedom of the performer can develop itself. The choice of classes may be different for different listeners and for the composer himself. It becomes evident that an "infinite music" is no more robust against casual or intentional fluctuations occurring in the performance. The gap between the acceptable freedom of any performer and the error tends to disappear, and not even the composer can fix the borderline. It turns out that continuous music is different in each performance, unless previously recorded.

In order to define a framework in which the roles of composers, performers and listeners are reasonably established, it seems to me that the uniqueness of the performance implies the identity of composer(s) and performer(s), which can take place only during an improvisation; in this way the musical performance becomes a happening, in which also listeners are involved as an active part, since also their reaction may influence the creation of the music.

REFERENCES


Fig. 1 - Free oscillations of a non linear oscillator.
From left to right: x vs t, v vs x.
a) undamped; initial velocity 0.2 (arbitrary units)
b) undamped; initial velocity 20.
c) damped with s=0.01; initial velocity 30.
Fig. 2 - Non linear oscillator with single sinusoidal force. 
From top to bottom: x vs t, v vs x, Fourier amplitudes of the stimulus (bottom) and of the response (top).

Fig. 3a - Non linear oscillator with two sinusoidal forces. 
From top to bottom: x vs t, v vs x, Fourier amplitudes. 
Driving frequencies \( f_1 = 10 \), \( f_2 = 15 \)
Fig. 3b - Non linear oscillator with two sinusoidal forces.
From top to bottom: x vs t, v vs x, Fourier amplitudes.
Driving frequencies f1=12, f2=20

Fig. 3c - Non linear oscillator with two sinusoidal forces.
From top to bottom: x vs t, v vs x, Fourier amplitudes.
Driving frequencies f1=10, f2=11
Music Cognition: Pitch, Scales, and Tonality
The Development of Absolute Pitch
Christina S. Chin
University of California at Santa Cruz

This theoretical paper discusses the nature/nurture debate surrounding absolute pitch. It will address three questions related to the development of absolute pitch: 1) What is absolute pitch? 2) How and when does absolute pitch develop? 3) Who has absolute pitch?

What is absolute pitch?

Absolute pitch (AP), commonly known as “perfect pitch,” is the ability to identify or sing pitches “out of the blue,” i.e. without an external reference. Absolute pitch can also be defined as long-term memory for pitches. In contrast to absolute pitch, relative pitch (RP) is the ability to distinguish intervals between pitches, or how far one pitch is from another. Intervals are ratios of frequencies, e.g. a fifth consists of two pitches whose frequencies are in the ratio of 2:3. All musicians have RP. In order to sing, AP is not necessary; all that is needed is to start off on the proper note, and then rely on RP ability.

Absolute pitch is not an all-or-nothing phenomenon. “AP possessors” usually refers to precise AP possessors who can name over 90% of isolated pitches correctly, but there are also imprecise AP possessors who can name around 50-80% of isolated pitches correctly, quasi-AP possessors who can produce one absolute reference pitch and use RP to determine the rest, and “absolute key” possessors (Bachem, 1937; Miyazaki, 1988; Sergeant, 1969). All AP possessors can recognize pitches, while only some AP possessors can also sing pitches on demand (Spender, 1980).

AP involves categorization and long-term memory; AP is not a superior pitch discrimination ability (Siegel, 1974). When AP possessors name pitches, they are categorizing pitches, a process which in some ways is similar to categorizing phonemes (Siegel & Siegel, 1977; Burns & Campbell, 1994) and colors (Bornstein, 1985). AP is a form of long-term memory, not a verbal mnemonic strategy (Zatorre & Beckett, 1989). On simple auditory tasks, AP possessors may rely on their long-term memory for pitch instead of using their working memory, as shown by most event-related brain potential (ERP) studies (Klein, Coles & Donchin, 1984; Wayman, Frisina, Walton, Hantz & Crummer, 1992; Hantz, Crummer, Wayman, Walton & Frisina, 1992). AP possessors, compared to non-AP musicians and non-musicians, do not have a P300 waveform, a possible indicator of working memory (Donchin, 1981).

AP may have a neuroanatomical basis. Schlaug et al. (1995), using magnetic resonance imaging (MRI), examined the planum temporale (PT), part of auditory cortex found in both temporal lobes. Although the PT shows leftward asymmetry in most people generally (Geschwind & Levitsky, 1968), and does to some extent for this whole sample, the AP possessors’ PT lateralization to the left was significantly greater than that of the non-AP musicians or the non-musicians.
How and when does absolute pitch develop?

AP possessors seem to have naturally acquired their ability without any deliberate effort, and it seems to be enduring. There is only one case of someone losing/unlearning AP, the accompanist Gerald Moore (Crutchfield, 1990). It is possible for highly motivated adults to train imprecise AP in themselves (Meyer, 1899; Brady, 1970). For adults, learning an absolute reference may be a more feasible strategy than trying to attain AP for all pitch classes (Cuddy, 1968). In order to develop precise AP, though, it is necessary to take advantage of a critical period in childhood. Although there are several reports of successfully teaching preschool-aged children AP (see Takeuchi & Hulse, 1993), this feat is not necessarily easily replicated (Cohen & Baird, 1990). While some researchers believe there is a genetic condition that is necessary, but not sufficient, for the development of AP (Profita & Bidder, 1988), the most recent comprehensive review of research on AP supports a theory of early learning, with younger children being more likely to develop AP because they are predisposed to attend to the absolute rather than relative features of melodies (Sergeant & Roche, 1973), and information in domains other than music (Takeuchi & Hulse, 1993).

Can these differing viewpoints about the etiology of AP be reconciled? The explanation of AP in terms of heredity (Profita & Bidder, 1988) does not specify what an inherited potential ability to develop AP is, beyond a sensitivity to musical tones, nor does it give a reason for why AP is only acquired in childhood. The explanation of AP in terms of environment (Takeuchi & Hulse, 1993) specifies why there is a critical period in childhood, when the cognitive shift from attending absolute to relative features occurs in many domains, but does not give a reason for why all children who begin musical training at young enough ages do not necessarily develop AP. I speculate about a way to synthesize these two explanations of the development of AP. Heredity and environment can both affect the two necessary ingredients for the development of AP, which are music training, and the tendency pay special attention to the pitch of music. Although music training is an effect of the environment, music lessons are usually initiated by parents, and it could well be that more musical parents are more likely to value music and ensure that their children have music lessons.

While the effect of music training on the development of AP is undisputed, the tendency to pay special attention to the pitch of music is probably representative of a more general tendency to focus on one dimension of music, or anything else, at a time. The tendency to focus on music unidimensionally can come from two sources: inherited cognitive style, and young developmental age. The unidimensional cognitive style variously called reflective, field-independent, left-hemispheric analytical, or narrow categorization (Zelniker & Jeffrey, 1979; Block, Gjerde & Block, 1986) is the cognitive preference to sequentially attend to one aspect of stimuli at a time, whether in music or in other domains. Since cognitive style is an individual difference that is relatively stable over the lifespan, it is probably inherited; a minority of people are classified in the category that is theoretically predicted to facilitate the development of AP. Interestingly, Bachem (1940) found AP to be more prevalent in the blind than in the sighted. This is probably because blindness results in the necessity of being sensitive to individual environmental sounds, which may make it natural to focus on individual musical pitches.
Young developmental age means being cognitively naive enough to not yet have shifted from thinking unidimensionally to multidimensionally (Siegler, 1996), from focusing on absolute to relative features (Takeuchi & Hulse, 1993), or from pre-conservation to conservation thinking (Piaget, 1950), in music and in other domains. Piaget (1950) originally specified that this transition from preoperational thought to concrete operations takes place between the ages of 7 and 8, but this position has been modified by later research to between the ages of 5 and 7 (Sameroff & Haith, 1996; Case, Okamoto et al., 1996). In a survey of over 1000 musically-educated college students and adults, Sergeant (1969) found that most musicians who started studying music before 7 years of age had AP, but most of those who started when older than 7 did not. In a study of 22 AP possessors, Miyazaki (1988) found that all precise AP possessors, and nearly all imprecise AP possessors, had started piano lessons between the ages of 3 and 5. In a study of 19 AP possessors, 25% of AP possessors realized they had AP by the age of 5, while 90% of AP possessors realized they had AP by the age of 10 (Profita & Bidder, 1988).

Music is just one domain of many in which children's thinking undergoes major change between the ages of 5 and 7. In music, children progress from understanding only first-order relations, i.e. associating a name with each pitch, to understanding second-order relations, i.e. associating a name with each interval, which is associating two pitches with each other. Evidence that children at these ages have achieved musical conservation is newfound ability to perform certain tasks, such as maintaining key when singing (Scott-Kassner, 1992) and recognizing transposed melodies (Bartlett & Dowling, 1980). In non-music domains, older children are able to reason much more sophisticatedly about such diverse concepts as letters, numbers, narrative, and space. This transition to "integrating relational premise information" (Halford, 1984) has also been referred to as "progressing from formal representation to symbolic representation" (Bialystok, 1992), or "merging two conceptual structures" (Case, Okamoto et al., 1996).

Musical savants are an exception to the cognitive shift taking place between 5 and 7 years of age. Musical savants are idiot savants especially talented in music, i.e. people with a low IQ or mental age, but musical ability corresponding to a much higher cognitive level. All 13 case studies of musical savants reviewed by Miller (1989) played the piano and possessed AP. It may be necessary for these low IQ individuals to possess AP in order to overcome their general cognitive limitations and develop musically. Both young children and low IQ individuals have limited working memories; AP, which is a long-term memory for pitches, can give these people an advantage in processing one aspect of music, pitch.

Who has absolute pitch?

Ward & Burns (1982) reviewed four types of theories about absolute pitch: genetic, learning, unlearning, and imprinting. Genetic/imprinting (Profita & Bidder, 1988) and learning/imprinting (Takeuchi & Hulse, 1993) approaches have already been discussed. The new way I am proposing to account for the data at hand is through a synthesis of the two positions into a genetic/learning/imprinting explanation. Research on absolute pitch has not come to a consensus about how often (or rarely) it occurs. Based on research in North America and Western Europe, absolute pitch is a very rare ability in the general population,
with estimates ranging from less than .01% (Bachem, 1940) to less than .07% (Profita & Bidder, 1988); among musicians, AP is still somewhat rare, with estimates ranging from 3.4% to 8.8% (Bachem, 1940; Miyazaki, 1988). Reportedly, however, in Japan as much as 30% (Kendall, 1996) to 50% (Miyazaki, 1988) of musicians may have AP, which is probably an effect of the widespread practice of starting children on music lessons at very young ages, and music education requirements in school (Lehman, 1993). For example, Miyazaki's (1988) study of college-level music and non-music students in Japan recruited, without trying, 40% subjects with AP. In contrast, Profita and Bidder (1988) were unable, in the U. S., to recruit two AP possessors married to each other.

Sergeant's (1969) finding that over half of the Canadian musicians who started music training before the age of 7 developed absolute pitch must be interpreted with the caveat that the majority of his sample were music professionals and music students. While people who choose to pursue music as a career are more likely to have started music early, which correlates with AP, by definition they have continued with music, making the discovery that they have AP more likely; it may be that the possession of AP helps motivate children to be seriously involved with music. To calculate an accurate percentage of AP possessors among people who studied music as children, let alone people who studied music before the ages of 5 to 7, it would be necessary to include all of the “serious amateurs” and “musical dropouts,” a task made difficult by high attrition in the study of music; in the U. S., two studies show 46% to 55% of junior high and high school students stop instrumental music before graduation (Weerts, 1992).

How can the different findings from research in the U. S. and in Japan be reconciled? In both places, not every child who starts learning music between the ages of 3 and 5 develops AP. The genetic/learning/imprinting explanation I have outlined in this paper suggests that not everyone can develop absolute pitch, but that more people than previously thought have the potential to develop absolute pitch. About 40% of people possess the cognitive style, or attentional tendency, to focus on one dimension of information at a time (Zelniker & Jeffrey, 1979), so about 40% of people have the potential to develop AP. In the U. S., however, music lessons are far from universal (National Assessment of Educational Progress, 1981), and for those fortunate enough to have had music training, it is probably safe to say that very few music students have started at the preschool level. In a survey of 2000 independent music teachers in the U. S., less than 1% taught preschoolers or adults (Uszler, 1996). In Japan, where the Suzuki method of teaching music is common, many preschool-aged children take music lessons. In 1964, upon Suzuki's first visit to the U. S., an American string teacher was surprised to discover that 4- and 5-year-old Japanese children were already playing relatively advanced repertoire, compared to 9- and 10-year-old American children beginning to study the violin (Mark, 1986).

Absolute pitch carries both benefits and drawbacks, depending on the musical context. Although AP can be an advantage when memorizing music or performing atonal 20th century music, AP can actually be a disadvantage when transposing music or working under mistuned conditions (Bachem, 1955; Crutchfield, 1990; Miyazaki, 1995). The term “perfect pitch,” and the scarcity of absolute pitch, imply that it is a special gift. If we give the gift of music lessons to more young children, however, we will find that absolute pitch is no longer so rare and mysterious.

108
References


To have it or not to have it, absolute pitch is the question.

Maite Moreno, McGill University
and
Valérie Descombes, McGill University

Introduction
Absolute pitch (AP), also known as perfect pitch (Ward and Burns, 1982), has captured the interest of many researchers. AP possessors are believed to be quite rare in the general population, though statistics vary with authors. According to Bachem (1955) the prevalence rate is 1:1000, whereas Profita and Bidder (1988) claim a rate of 1:1500 or less. Wallin (1985) cited that 87% of a group of highly qualified performers possessed AP. Why some people have AP and others do not has been a matter of some controversy. Many of the studies on AP have focused on this nature-nurture problem. While some researchers considered AP to be inborn or hereditary (Bachem, 1950; Révész, 1953; Seashore, 1938), others support the theory that it can be acquired through intensive training when starting at a young age (Crozier, Robinson et Ewing, 1977; Sergeant, 1969; Miyazaki, 1988; Ogawa and Miyazaki, 1994), or at any age (Brady, 1970; Cuddy, 1970; Lundin, 1963; Rush, 1989).

This research has greatly contributed to our understanding of AP. However, in whatever manner AP is acquired, the practical question remains, is there an advantage in possessing this ability? Some authors share their opinions by suggesting that it is an advantage for sightreading and in musical dictation when faced with difficult passages (Miyazaki, 1992). Others signal its use in the performance of a capella singing (Ward and Burns, 1982), “playing an instrument in tune and knowing what a piece of music sounds like simply by reading the score” (Eaton and Siegel, 1976). There also would seem to be quite a few disadvantages such as problems with music tuned lower than usual, as in Baroque music, (Miyazaki, 1992), or problems with transposition (Miyazaki, 1992). In their studies Burns and Ward (1974, 1978) found possessors relying solely on AP and thus having difficulty in relative pitch tasks such as interval categorization. Bachem (1955) also found possessors making octave errors, which he called “chroma pitch” errors. In summary, there are contradictory positions to the real value of having AP. Some say AP is of doubtful value to a musician (Chailley, 1985; Ward and Burns, 1982), while others assert that it is an extremely important musical ability (Steblin, 1987). In order to determine the practical importance of AP we are interested in confirming these opinions with more extensive research.

Purpose
The main purpose of this study was to investigate the advantages and disadvantages of having AP. The nature of these advantages and disadvantages will also be examined.

Method
Procedure
All data was collected through a WWW site on McGill University's Graduate Student Research homepage. The implementation of the homepage followed the model developed at McGill's music department (http://lecaine.music.mcgill.ca/~welch/auditory/Auditory.html).
Subjects
People from around the world who connected to the World Wide Web (WWW) were the subjects in the project. The site was opened to the public on April 16th, 1996, and the last data were collected on July 3rd, 1996 (a total of 78 days). To increase the quantity of subjects in a relatively short period of time and to target populations that may have strong opinions on the use of AP, advertising of the project was done through e-mail directories.

Questionnaire
The data was collected via a questionnaire comprising of three sections: (1) background information; (2) opinions on the advantages and disadvantages of having AP; (3) an auditory test measuring AP ability.

The setting of the questionnaire was designed in two layers. The first layer contained introductory material, the three sections of the questionnaire as mentioned above, and hypertext links. At the end of this page the subject had access to the second layer, the audio test, by clicking the submit button. The participant was presented with instructions and five loudspeaker icons, where they were free to sample the sounds in any order. When an icon was accessed, the server randomly chose a pitch from C₄ to B₄. All audio examples were pure monophonic sine tones synthesized at a 22,050 sample rate with 16 bits per sample. Two formats were given in option to the user to transfer the audio data: WAV, suitable for most PC systems and AU, suitable for Macintosh, Sun, and NeXT computers.

The questionnaire (text and pitches) was tested to ensure proper formatting and quality. To ensure correct pitch height of the pitches when accessed from different machines, two AP possessors took the pitch identification test each on two different machines. All four test results were graded 5/5, thus the pitches were deemed acceptable.

The pitches for the audio test were coded so that subjects would not be able to decipher the pitch names when downloading them. The subjects, however, were able to receive the pitch identification results by entering their e-mail address. After submitting their answers, the results, comprising of the pitches heard and the subjects' answers, were automatically e-mailed to them. All data were automatically gathered by the server. The data for the questionnaire and the pitch identification test were gathered on two separate files. Each downloading was date stamped and identified the clients machine. This facilitated the investigators task of coding and retracing each file. This also enabled the investigator to match the questionnaire with the audio test, number the subjects, and exclude repeated submissions.

Results
Table 1.

<table>
<thead>
<tr>
<th>Gender:</th>
<th>Male (125); Female (111); unknown (5).</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Age:</th>
<th>0-17 (28); 18-22 (57); 23-30 (57); 31-50 (85); over 50 (10); unknown (4).</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Have AP:</th>
<th>yes (52); no (129); maybe (60).</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Would like to have AP:</th>
<th>yes (184); no (20); maybe (34); unknown (3).</th>
</tr>
</thead>
</table>
Of the 248 submitted questionnaires, 241 were used in the analysis. The remaining seven were left blank and were therefore unusable. The demographic characteristics of the 241 subjects used in the study are presented in Table 1.

The main purpose of the study was to determine the advantages and disadvantages of having AP. The subjects' comments (from both possessors and nonpossessors) were categorized by the investigator according to the most frequent statements. To check for the reliability of the classification of these categories, the comments of 10 subjects were reclassified by a second judge. The percent agreement was 96%.

Thirteen categories were created for the advantages, yielding a total of 277 comments (Figure 1). The advantage of tuning an instrument or singing a pitch without a reference was cited 60 times (22% of the total advantages). Continuous pitch accuracy while playing an instrument or singing was reported 53 times (19%). AP was said to be advantageous for sightsinging and sightreading by 30 subjects (25 and 5 respectively, totaling 11%), as well as for improvisation and playing by ear (11%). Transcription was explained as being easier and faster when one had AP by 22 subjects (8%). Impressing your
friends with AP qualities was reported 17 times (6%). The ability of identifying the key of a piece of music and hearing its harmonic structure was also counted 17 times 6%. The advantage of knowing what notes are heard, such as acute aural perception, ability to express pitches in Hertz, and identifying non-musical pitches, was reported 15 times (5%). Fourteen people said AP aided in composing, 9 in memorizing, 4 in hearing the musical score in ones head, 3 in visualizing the notes that one hears, and 3 increasing ones authority and credibility.

Nine categories were created for the disadvantages, yielding a total of 156 comments (Figure 2). The annoyance of hearing inexact tuning received the highest frequency rate with 73 citations (47%). Problems with transposition, either playing a transposing instrument, playing or singing a piece in another key than is written, or listening to a piece played in a different key than is usually played in, were mentioned 30 times (19%). Tuning away from A=440), such as Baroque music, was mentioned 17 times as being possible disturbance by possessors (11%). Eleven enumerated possible over analyzation by possessors (7%). hearing each note in a piece of music rather than its entirety (the overall expression) was cited as a disadvantage 8 times (5%). Relying too heavily on the AP ability and therefore not being able to identify intervals or chords, as well as more is expected of you was stated 7 times (4%). One percent said that AP hindered creativity and one person mentioned that AP gets "out of tune" with age.

Subjects usually gave one to two comments for each advantage and none or one disadvantage. Overall, more advantages were cited than disadvantages (Figure3), regardless of possession (Figure 4).

Of the 241 subjects who responded to the questionnaire, 69 also completed the audio test. 172 were either left blank or unsubmitted, leaving 60% of the pitch identification tests unavailable for
further analysis. All subjects who took the test and claimed "maybe" having AP received a score of 3 or lower (out of a maximum score of 5) on the pitch identification test. These subjects were classified as nonpossessors (Table 2).

### Table 2. Subjects' AP test scores

<table>
<thead>
<tr>
<th>Have AP</th>
<th>AP pitch identification score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>No</td>
<td>20</td>
</tr>
<tr>
<td>Yes</td>
<td>3</td>
</tr>
<tr>
<td>Maybe</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>31</td>
</tr>
</tbody>
</table>

**Discussion**

The results of this study support previous observations on the advantages and disadvantages in relation to the AP ability. Some prevalent advantages in relation to AP are in such uses as in sightreading, music dictations, playing in tune, improvising, and composing (Miyazaki, 1992; Profita and Bidder, 1988). In the present study the use of AP in such instances was reported by quite a number of people. The disadvantages of such problems as with music tuned lower than usual (as in Baroque music), problems with transposition, or relying solely on AP and thus having difficulty in relative pitch tasks such as interval categorization (Ward and Burns, 1982) were also comments of the subjects in this study. However, it seems that people are more preoccupied with the advantages and disadvantages of AP in relation to intonation. Furthermore, most possessors answered affirmatively to the question of desire in having AP, suggesting that even in the face of these liabilities, it is not easy to find anyone willing to give it up.

One subject mentioned the disadvantage that AP gets "out of tune" with age. Though only one subject spoke of this, the comment was reported due to its interest. There are reports in the literature that many individuals with AP experience a shifting of their sense some point around midlife: the pitch "sharpens" by as much as a semitone or even more (see, e.g., Ward and Burns, 1982). It would seem that a certain amount of regular exposure to music is necessary in order to maintain AP at its highest level of functioning. Though no relation was found with possession and the number of years that a subject played an instrument, this brings us to another issue, that of the definition of AP.

Many subjects in this study reported "maybe" having AP. Incidence of AP may be difficult to determine primarily because of problems of definition, identification, and testing. Rush (1989) points out that the term "perfect pitch" raises some important semantic problems, for people can make certain kinds of mistakes and still be said to have "perfect" pitch. He suggests that the term "absolute" at least draws attention to the labelling of pitches on an absolute, rather than referential, basis. The identification of unusual sensitivity to musical tones is complicated by the fact that individuals possess the ability to varying degrees or in special ways (Bachem, 1937; Rush, 1989). It is not known whether musical tone sensitivity or perceptiveness is a skill or aptitude which varies smoothly or continuously between these extremes, or whether it is a unique single trait. Testing therefore is also problematic. Depending on the varying degree of AP
ability, or the influences of its acquisition, testing becomes a delicate situation. Perhaps many individuals with innate AP are never discovered simply because they never encounter conditions whereby it can be identified.

In this study technology played a large role in the testing. Many subjects explained inability to download the pitches due to their systems. Furthermore, the quality of the pitches is not assured to be accurate.

Nonetheless, the results of the study confirm that AP possessors are not amongst the majority of the people and that even the ones who claim having this trait, few actually possess absolute pitch.

It is clear however that both AP possessors and nonpossessors found this ability to be more advantageous than disadvantageous, therefore research efforts should persist to confirm the reality of these opinions.

References
Pitch memory and the A440 Scale

Daniel J. Levitin, Interval Research Corporation
Palo Alto, California U.S.A.

Abstract. Previous studies have shown that people have accurate long-term memory for perceptual events. The present experiments take this notion further, providing evidence for "high-fidelity" memories for musical pitch. The present study shows that normal subjects, unselected for musical ability, appear to have assimilated the markers (or "focal tones") of the Western scale into long-term memory. Although subjects were unable to demonstrate this in a vocal production task, when asked to adjust a variable sine tone oscillator to imagined tones, 24% came within 5 cents of the nearest focal tone. The relevance to theories of absolute pitch and musical memory is discussed.

Introduction. A casual reading of the memory literature across the last few decades gives one the certain impression that most memory experimentalists are primarily concerned with what people forget, and that relatively little attention is paid to what it is people actually remember. Fortunately, a number of studies have recently begun to focus on how well we remember, and with what degree of precision. A primary example in the visual domain is the work of Magnussen and Dymes (1994). Their subjects were shown photographs of sine-wave gratings for discrimination at various retention intervals. Performance following a two day retention interval was equivalent to performance following a one second retention interval, demonstrating visual discrimination abilities in the hyperacuity range even when these discriminations are made from memory.

Following the lead of a number of researchers (e.g. Deutsch, 1991; Lockhead & Byrd, 1981; Terhardt & Ward, 1982) I conducted a series of experiments demonstrating that details of auditory experiences are recorded in long-term memory with great accuracy. In particular, subjects asked to sing their favorite popular music songs tended to do so in the correct key and tempo (Levitin, 1994; Levitin, 1996; Levitin & Cook, 1996). I argued that repeated exposure creates a representation of the song in long-term memory that preserves accurate pitch and tempo information. This is contrary to those memory theories that propose that the function of memory is to abstract the gist of experiences at the expense of specific details. It is consistent however with those theories that postulate the storage of multiple-traces and specific instances (such as Bower, 1967; Hintzman, 1986).

Given that people remember the actual pitches of their favorite songs, a reasonable question to ask is "with what level of precision?" In other words, how close to the canonical pitch do subjects come when remembering musical tones? One could argue that a lifetime of exposure to a particular tonal system (for example, the present-day A 440 Hz standard) would leave a series of traces in memory, forming a sort of template, with markers along the frequency continuum denoting the location of each chromatic tone. (For example, memory markers at 440 Hz, 466.16 Hz, 493.88 Hz, etc., corresponding to A₄, B♭₄ and B₄.) If this were so, subjects in a production task should tend to produce tones that cluster around the chromatic scale tones (or "focal" tones) of Western music with a more or less mound-shaped distribution. On the other hand, if people do not have an accurate memory for the location of pitch markers, we would expect their productions to be distributed more or less uniformly around the focal tones. Although the locations of focal tones along the frequency continuum are well-defined by convention, these locations are arbitrary, and frequency space is continuous, not discrete. Therefore, a reasonable null hypothesis (if people have not assimilated information about focal tones into long-term memory) is to find a uniform distribution of productions along the frequency continuum.
In the present experiments, it is important to distinguish between the contents of memory and a subject’s ability to match those memories in vocal performance. That is, failure to find a mound-shaped distribution in a singing task does not necessarily mean that the mental representation of pitch lacks precision, because poor vocal control could cause such a result. An alternative to singing would be to ask subjects to set a variable tone generator to match a pitch they imagine. Therefore, these experiments employed both production methods, singing and the setting of a variable tone oscillator.

Procedure. Subjects were asked to think of a random tone and then to sing this tone using the syllable “la,” or to try to produce it using a variable sine tone generator (depending on condition in a between-subjects design).

Results. The research question was whether people producing tones would tend do so at or very near the focal pitches in the Western A440 scale. In analyzing the productions, pitch class was ignored, and only cents deviation from the nearest focal tone was examined. The data for the singing and oscillator adjustment conditions are presented in Figures 1(a) and (b) respectively. These figures show the data plotted as histograms around cents deviation from the nearest focal tone, in ten cent-wide bins. Because -50 cents and +50 cents are equivalent, only the +50 cents bin is shown. The data are presented as circular histograms in order to graphically convey the equivalence of the ±50 cent categories; such circular representations are the most appropriate way to view these types of data (Fisher, 1993).

![Figure 1a. Distribution of errors for singing a random tone.](image-url)
For the singing condition, Kuiper's $\gamma$ test of uniformity indicates a good fit with a uniform distribution ($U=0.36$, n.s.), suggesting that people do not sing tones near the markers of the western scale in this task. However, the data for subjects who adjusted a tone generator describe a normal distribution with 0 cents deviation as the mode, and the the $\gamma$ test leads us to reject the hypothesis of uniformity in favor of unimodality ($U=1.98$, $p<.02$). This result supports the hypothesis that subjects have high-fidelity memory for the location of focal tones in the Western scale, and that despite an inability to produce these tones vocally, they are able to imagine them and match them with a variable oscillator. Note that 24% of the subjects in the adjustment condition (9 out of 37) came within ±5 cents of a focal tone, and 46% of the subjects came within ±15 cents of a focal tone. By chance, one would expect only 10% of the subjects to come within ±5 cents, and only 30% to come within ±15 cents.

Another measure of performance is the absolute value of the average production error. (This absolute value measure is the appropriate measure of errors, because subjects could make errors in each direction from a focal tone, and thus the algebraic means would confound systematic and random sources of error.) The absolute value of the average error for the tone oscillator subjects is 19 cents. By comparison, the absolute value of the average error for traditionally-defined AP possessors using an oscillator to produce pitches has been variously reported to be 11 cents (Bachem, 1940), 17 cents (Siegel, 1972), and 25 cents (Levitin, 1996). The present subjects - unselected for musical training - are thus performing well within the range of what we might think of as "traditional absolute pitch" performance.

Discussion. The evidence from this experiment suggests that long-term memory does indeed preserve a high-fidelity representation of the focal pitches in the Western scale. This was revealed in a pitch production task using a variable oscillator, in which subjects tended to produce tones that clustered around the focal tones, with a more or less mound-shaped distribution. The findings can be interpreted as evidence that normal subjects, unselected for musical ability, have something akin to absolute pitch (AP). By definition, AP is the ability to identify or produce tones from memory. Historically, AP has been tested by providing the subject with a label (such as C#, 440, or Do) and asking the subject to produce the corresponding tone, or by playing the subject a tone and asking them to provide the associated label.
In a previous paper (Levitin, 1994) I argued that AP actually consists of two separable component abilities, pitch memory and pitch labeling. I introduced a method for testing pitch memory by asking subjects to sing their favorite popular music songs, and comparing their productions with the actual recordings. The subjects in that study were found to have accurate pitch memory even in the absence of pitch labeling, leading me to speculate that a much larger proportion of people have latent AP than has been previously recognized. The subjects in the present study have shown that they have assimilated the markers of the Western scale into long term memory, and that they are able to produce them using a variable oscillator. This can be seen as further evidence for the widespread prevalence of a form of AP, pitch memory.

It is important to reiterate that the actual location of the focal tones is arbitrary both from a music theoretic and physiological point of view (indeed, the A440 anchor has become standard only in the last 80 years or so). Therefore, knowledge of these markers must have been learned through repeated exposure to them. Perhaps all that distinguishes "true" AP possessors from these "latent" AP possessors is that "true" AP possessors have learned to associate verbal labels with their internal templates (Levitin, 1996). Recent findings that pitch perception in AP subjects can be altered by administration of certain psychoactive drugs (Chaloupka, Mitchell & Muirhead, 1994) lend further support to the internal template hypothesis of absolute pitch.

In reference to my earlier study (Levitin, 1994), some critics have argued that the subjects' ability to sing the correct tones of their favorite songs did not indicate pitch memory, but rather, that it merely indicated good muscle memory for vocal chord tension. Of course, this argument overlooks the fact that muscle memory is a form of memory, and the ability to associate precise muscle states with particular pitches seems quite remarkable to me. Nevertheless, a key study by Ward and Burns (1978) showed that subjects who were denied auditory feedback, and thus had to rely solely on muscle memory, did not do a very good job of matching a pitch: many made errors as great as a third of an octave. However, the present finding that untrained singers were capable of adjusting a tone generator with great precision to match an (alleged) pitch template in long-term memory, provides strong evidence that internal representations for auditory perceptions are accurate even when decoupled from kinesthetic memories.
References


Author's Note. I am grateful to two undergraduate research assistants, Anne Stern and Daniel A. Stern (no relation), for collecting some of the data reported herein. Some subjects were paid for their participation from ONR Research Grant N-00014089-3013 to the University of Oregon. Address correspondence to levitin@interval.com.
The octave, a fundamental concept of Western music, is yet to be fully understood. In informal discourse the relation between two tones a number of octaves apart is often construed either as strict identity, or else as watered-down "similarity." In semi-formal and even formal discourse "octave equivalence" is often believed, erroneously, to imply inversive equivalence of intervals (e.g., P4=P5).

I shall begin by making a formal distinction between two somewhat different octave concepts, "octave equivalence" and "octave relatedness." The proposed distinction is essentially an algebraic one, and concerns the types of objects with which one chooses, a priori, to work. Suppose one chooses to work with sets of equivalence classes, say, the set of seven diatonic note-classes \{C, D, \ldots, B\} (and their intervals). Although any given note-class (say, "C") consists of register-specific notes (an indefinitely large number of such notes, conceptually at least), the individuality of each such note is lost within the class as a whole. In other words, in working with note classes one cannot refer to a particular member of a given class (say, "C4") except as a representative of the class as a whole (thus C4 represents C: the class of all C's). Geometrically, octave equivalence in this sense may be captured by a circular arrangement of note-classes, as in Fig. 1.

![Figure 1. Octave equivalence: A circle of note classes](image)

Alternatively, one might choose to work with a set of register-specific notes (and their intervals), all subject to an "octave relationship" such that, say, C4 and C3 are in that relationship. This sense of octave relatedness may be captured geometrically by a helical arrangement of notes, as in Fig. 2. Although the so-called "helical model of pitch" is well-known, it is worth emphasizing that this model, unlike the circular model, is register specific. That is to say, the helical model is register-specific, while at the same time embodying an equivalence relationship—the octave—by which any given note (say, C4) relates to other register-specific notes (thus C4 is octave related, for example, to C3); indeed, the set of all notes to which a given note is octave related is identical to an equivalence class in the "circular" sense.
It should be clear that octave relatedness is the more inclusive octave concept, for the helix contains the circle, but not vice versa. Special care must be taken, therefore, in assuming that octave equivalence rather than octave relatedness applies in a given context. Indeed, William Benjamin (1974, p. 172) rightfully questions whether Forte’s axiomatization of octave equivalence in *The Structure of Atonal Music* (1973, p. 2) is not “a trifle glib.” Diana Deutsch conducted many experiments that test the processing of octave relationships (e.g., Deutsch 1972, 1973, 1976, 1978, 1983; Deutsch & Boulanger 1984). Deutsch believes, even in the face of seemingly conflicting evidence obtained by other researchers (e.g. Idson & Massaro 1978; Kallman & Massaro 1979), that her experiments support her (1969, 1982) “two-channel” model for music recognition, a model where octave equivalence is assumed in the context of pitch-simultaneities (chords), but not pitch successions (melodies). “The results of the present experiment” write Deutsch and Boulanger (1984, pp. 47-8) “are in accordance with the two-channel model for the abstraction of pitch relationships, from which it is predicted that octave equivalence effects should not operate directly in the processing of melodic intervals.”

I believe that Deutsch and her collaborators are quite correct in separating the melodic domain from the chordal domain in this context; nonetheless, they seem to overstate their claim. Although the assumption of octave equivalence (a circle of note-classes) is problematic from the melodic point of view (but not from the chordal point of view), the assumption of octave relatedness (a helix of notes), by comparison, is (melodically) indispensable. Indeed, extensive experimentation is hardly necessary to establish the significance of the octave relationship in the melodic domain. To be sure, assumptions must always be questioned; and yet, I find it hard to believe that one could seriously question the assertion that, in Ex. 1, the asterisked notes are related in a very special way.

Example 1. Schubert, “Wasserflut” (*Winterreise*, No. 6), mm. 5-8
Interestingly, Deutsch’s “neural-net”-type model implements an idea closely related to David Lewin’s (1987) algebraic notion of *Generalized Interval System* (GIS). A GIS consists of a set $S$ of “pitches,” a group IVLS of “intervals,” and a function “int” that maps all possible ordered pitch-pairs into IVLS. Fig. 3, redrawn on the basis of both Deutsch 1969, Fig. 5, and Deutsch 1982, Fig. 2, represents the first of two “channels” in Deutsch’s model. Deutsch’s “primary array” is, essentially, a set of notes, and is thus analogous to Lewin’s set $S$. The elements in Deutsch’s “Stage 2” are, for the most part, intervals, and are thus analogous to Lewin’s group IVLS (Deutsch, however, also includes triads at this stage). Finally, Deutsch’s “Stage 1” connects note-pairs (for the most part) from the primary array with the intervals of Stage 2; thus, this stage performs analogously to Lewin’s function “int.” As Lewin has shown, it is possible to construct a GIS with either (1) pitches and intervals; or (2) classes of pitches and intervals modulo the octave. The former GIS, together with the octave relationship, corresponds to the helical model (octave relatedness); the latter GIS corresponds to the circular model (octave equivalence).

While the assumption of octave equivalence (a circle of note classes) is problematic from the melodic point of view, from the chordal point of view the situation is quite different. Unlike melodies, chords do seem to retain their identities under octave displacement of their component notes. Consider, as an example, the chords $K$ and $L$ in Ex. 2. From the melodic point of view the downward octave displacement of the note G in the upper voice is extremely important; yet from the chordal point of view this displacement may be practically ignored. The situation is slightly more complex with chord $M$. On the one hand, $M$ (in Roman-numeral analysis, a “1” in C major) is the same as $K$; yet on the other hand $M$ is an “inversion” of $K$.

In what sense is chord $M=\{G,C,E\}$ “the same” as chord $K=\{C,E,G\}$? According to Diana Deutsch (1982, p. 273), “given the principles of octave equivalence and interval equivalence,
one might hypothesize that intervals whose components are placed in different octaves are also perceptually equivalent. This assumption is frequently made by contemporary music theorists who refer to such intervals as in the same interval class." Indeed, Peter Castine believes that

Since pitch classes embody octave equivalence, there is no way to determine which of two pitch classes is 'higher' than the other. To make the point using traditional notation, consider the notes D₄ and A₅. If we ignore register, there is no way to determine whether the interval is that of a perfect fourth or a perfect fifth. So, when dealing with pitch class, we must consider interval inversions to be equivalent. (1994, p. 29)

Yet contrary to popular belief, octave equivalence does not imply that inversionally related intervals are equivalent. In the group \( \mathbb{Z}_{12} \) of integers modulo 12 -5 is in the same equivalence class as 7, but not in the same class as 5 (the inverse of 7). Similarly, in the group of "note-class intervals" the descending perfect fourth is in the same class as the ascending perfect fifth, but not in the same class as the ascending perfect fourth. The question of so-called "inversional equivalence" of intervals (e.g., \( P_4 = P_5 \)) does not concern octave equivalence at all, but rather the notion of "distance" (as distinct from "interval"). Indeed, there exist strong algebraic arguments for not taking inversionally-related note-class intervals as (distance) equivalent, unlike "simple" note-intervals (Agnon, 1996).

The equivalence of \{G,C,E\} and \{C,E,G\}, therefore, is not due to the "equivalence," say, of the perfect fourth G-C and the perfect fifth C-G (the note-class interval "perfect fourth" is simply not equivalent to the note-class interval "perfect fifth"); rather, the two chords are equivalent because Roman-numeral analysis assumes that chords are ordered in a specific way, namely root-third-fifth (or root-third-fifth-seventh in the case of seventh chords). Thus, if we have any fourth at all in \{G,C,E\} it is a descending one, from C to G (root to fifth); this descending fourth is of course in the same class as the ascending fifth.

We now turn to the sense in which \{G,C,E\} is distinct from \{C,E,G\} ("I₁²" as opposed to "I₂¹"). The two chords are distinct in the context of a different method of chordal analysis known as figured bass. In figured-bass analysis chordal note classes are partially ordered (Morris, 1995); unlike Roman-numeral analysis, however, the "first" note class of any chord is not the root, but rather the bass. Thus, in \{G,C,E\} we indeed have a fourth G-C, as distinct from the fifth C-G.

Here, however, a difficulty arises. If, indeed, one is working with note classes rather than notes, how can a bass exist? In other words, how can the notion of a "lowest tone" coexist with one's a priori decision not to distinguish, in the chordal context, between octave-related notes? (This difficulty does not arise in the Roman-numeral context due to the more abstract nature of the order root-third-fifth.)

The difficulty, however, is more apparent than real. In both the Roman-numeral and the figured-bass contexts one is indeed concerned with note classes (octave equivalence) rather than notes (octave relatedness). Registral information in the figured-bass context does not concern the nature of the elements under consideration, but rather their order; registral information is available for this particular purpose since the bass, as a melodic entity, is composed of register-specific notes.

Ex. 3 may help clarify that register is indeed an "external" issue in this context. Suppose this excerpt from a Bach chorale is performed a cappella. How is chord X to be regarded in terms
of figured bass? Taking the lowest-sounding voice as “first” X is a second-inversion triad. This characterization, however, seems counterintuitive. A strong connection seems to exist between X’s D4 and the initial D3 in the bass (note the ascending octave leap), and similarly, between X’s A3 and the preceding B♭3 in the tenor (note the stepwise connection). In other words, it makes sense in this context to regard D4 as X’s conceptual “bass” (as notated by Bach), even though A3 is in fact the lowest sounding tone. Clearly, with D4 as the bass X becomes a root-position triad. In particular, the note-class interval from the “bass” D4 to the “inner voice” A3—a descending fourth—is in the same class as the ascending fifth.

Example 3. Bach, Chorale (“O Ewigkeit, du Donnerwort”), Riemenschneider No. 274, mm. 3-5

In summary, the concept of the octave varies according to context. In melodic contexts one has a helical arrangement of register-specific notes (octave relatedness), whereas in chordal contexts, a circular arrangement of note classes (octave equivalence). In chordal contexts, in addition, there are two main methods of ordering note classes for the purpose of calculating interval-content (in melodic contexts notes are ordered temporally). Roman-numeral chordal analysis is a method by which chordal note-classes are ordered as root-third-fifth(-seventh), whereas figured-bass chordal analysis is a method by which note classes are only partially ordered, the bass being “first.” Finally, contrary to popular belief octave equivalence does not imply “inversional equivalence” of intervals.
References


Scales, Sets, and Interval Cycles: A Taxonomy
John Clough and Nora Engebretsen
State University of New York at Buffalo, USA

Introduction. Recent studies in the theory of scales by Eytan Agmon (1989), Gerald Balzano (1980), Norman Carey with David Clampitt (1989), John Clough with Jack Douthett (1991), Clough with Gerald Myerson (1986), and Carlton Gamer (1967), which may appear diverse, have in common the central role of a generator—an interval whose repetition ties together all the pitch-classes (pcs) in a set or segment. The set or segment thus formed is commonly known as an interval cycle, either complete in the sense that additional repetitions of the interval yield redundant pcs, or incomplete in the sense that they do not. Drawing largely on properties defined in these studies, we propose a taxonomy for the special categories of pitch-class sets (pcs) that correspond, collectively, to the interval cycles (complete and incomplete). The taxonomy is, we believe, interesting in itself as it addresses complex interactions among features that theorists have defined for different purposes. We believe as well that it strengthens interrelationships among the three concepts of the title: ideas from the theory of scales are cast across pitch-class space, and as a result we discover anew the place of the interval cycle as a pervasive and unifying theoretical notion.

In addition to the common thread of the interval cycle, the cited studies also share a common motivating factor: the desire to explore spaces defined by features of the usual diatonic scale. If we regard a particular property of the diatonic scale as interesting, it is natural to search for other scales with that property, as a means of gaining insight about it. The special status of the generator (or interval cycle) as a feature of many classes of scales will become clear as our scheme unfolds.

We presume familiarity with 12-pc space; and we hope that our classification will be seen as meaningful for the music of that familiar space; however our results are intended to apply quite generally to microtonal systems and to all contexts where ordinal distances operate as, for example, in cycles of generic intervals—3ds, 4ths, etc.—within scales.

Our program is as follows: We first define eight features pertaining to scales, all but one of which has been previously defined in the literature. After observing logical relationships among the features, we identify twelve sets of features which serve to partition pcs sets that have at least one of the features. These twelve feature-sets support the proposed taxonomy, which amounts to a mapping of pcs sets onto the feature-sets. The range of individual features over the twelve feature-sets then comes into view, followed by algorithmic approaches to the exhaustive generation of sets with particular features or combinations of features. We close with remarks on a newly defined category of sets.

Features. Example 1 lists the scale features recognized in the present study, with brief definitions and one or two examples of pcs sets that exhibit each feature. In some cases we have substituted definitions equivalent to those in the cited sources, to suit our objectives. Pcs sets are enclosed in curly brackets; a following subscript indicates the size of the modular chromatic universe if other than 12. To simplify the exposition, and with no loss of precision, we dispense for the most part with the concept of set-class and deal only with literal sets: it is clear that if a pcs set has any of the defined features, then all members of the corresponding set-class (under the usual canon of transposition and inversion) have the same features.

With these preliminary notions in place, let us take a brief tour of the definitions of Example 1. G sets are quite simply those sets associated with interval cycles. We use Marc
Wooldridge's (1992) notation for such sets, where parameters are given as shown at the top of the center column in Example 1. For convenience, we assume throughout that pc 0 is the "origin." WF sets are as defined by Carey and Clampitt (1989); each generating interval spans a constant number of scale steps. (WF sets are, with the exception of total-chromatic sets [equal-tempered scales] precisely those defined earlier, in different terms, by Erv Wilson [1975].) MP scales, in which each generic interval comes in two specific sizes, were defined by Clough and Myerson (1985). DE scales, in which each generic interval comes in either one or two specific sizes, are isolated here for the first time, as far as we know. ME sets were defined by Clough and Douthett (1991). DT scales or systems, in the larger sense of the term, were defined by Agmon (1989); however we give here the equivalent definition provided by Clough and Douthett, as it best comports with other definitions used here. DP scales were defined by Gamer (1967), and, finally, BZ scales by Balzano (1980).

To be sure, a great many additional features of scales and sets have been defined in the literature. Those selected have arisen in the course of studies based on the remarkable set of features possessed by the usual diatonic. But they are by no means the only features to have arisen in this way. Indeed, a number of others deserve mention. Although we exclude them from our system of classification here in order to make the task more tractable, some or all of these additional features might well be included in refinements of the present system, or in classifications designed for other purposes. We will mention in passing five such features: inversional symmetry, non-trivial transpositional symmetry, prime cardinality, Agmon's (1991) efficient linear transformation (ELT) property, and that which we shall call Cohn's property, in recognition of Richard Cohn's (1996) work isolating the class of pc sets associated with what he terms maximally smooth cycles.

Feature sets. Having defined the several features that we shall be concerned with, we now look at how they combine. As illustrated in Example 2, a number of implicative relationships are present among the features listed in Ex. 1. For instance, all diatonic sets have Myhill's property, as shown by the arrow from property DT to property MP in Ex. 2. Further, by following all paths from property DT consistent with the arrows in Ex. 2, we see that diatonic sets necessarily have all the other properties except Balzano's property. In fact the diatonic set in 12-pc space is the only diatonic set that does have BZ and hence all of the eight features; it is unique in this respect. If a set has any of the listed properties, except ME or DE, it also has G. And so forth. Note that the implications shown are all one-way; which is to say that no two features are logically equivalent. The broken arrow from DE to G symbolizes a somewhat different relationship that we will discuss later.

There are sets of features consistent with Ex. 2 that, as far as we can tell, are nowhere instantiated by actual pc sets, although we cannot prove it. We are in fact able to construct pc sets consistent with just twelve of the twenty sets of features that are consistent with Ex. 2. The peculiarities of instantiation with respect to Ex. 2 seem to arise from the fact that the various features address properties that are not all commensurable with one another.

Let us now look at sets of features that are instantiated by pc sets. They are listed in Example 3, along with a sample pc set corresponding to each F-set. The checks and apples in each row mark the features of the corresponding F-set; apples mark a subset of an F-set that implies the entire F-set, based on the information of Example 2. We conceive the relationship between the pc sets and the F-sets to be a many-to-one mapping from the former onto the latter: a given pc set corresponds to one and only one F-set—the F-set with precisely all of its features. As we have not proved that all pc sets in all chromatic worlds are accounted for by this list of F-sets (to which may be adjoined the null subset of the features), it is conceivable that additional instantiated F-sets remain to be discovered.
There are complex relationships within and among conglomerates of sets defined by means of logical operators over the features. Suppose, for example, that we wish to look at pcsets that have DE and not WF. We see that such pcsets correspond to F-set 6 (DE only) and F-set 7 (ME and DE). Given that the hexatonic and the octatonic are examples, respectively, of these two F-sets, we may reasonably guess that the pcsets consistent with the logical expression DE and not WF are precisely those that are symmetrical under (non-trivial) transposition and symmetrical under inversion (Messiaen’s modes, for example), some of which are ME and others not. This is in fact the case.

The above conglomerate of pcsets, defined by a simple logical expression over the features, and corresponding to just two F-sets, is relatively easy to intuit. As an aid in grasping larger conjunctions of F-sets that correspond to more complex logical expressions, Venn diagrams are useful.

**Enumeration.** It is one thing to characterize a family of pcsets in terms of a particular F-set and another to show how the family’s pcsets may be exhaustively enumerated. Since it is possible to devise an algorithm to enumerate all pcsets in all chromatic universes, we need merely adjoin an appropriate sorting function to such an algorithm in order to enumerate the pcsets corresponding to any or all F-sets. Sadly, the design of such a procedure would confer little or no additional insight—although observing the sequence of its sorted output could be quite interesting. For the twelve F-sets, however, we have developed algorithms based on implied secondary features not explicitly part of the feature set—algorithms that do yield insights as to the proclivities of the defined families of pcsets.

The basic methodology underlying the development of these algorithms was to begin with a “filtering” algorithm—essentially a sieve designed to capture desired properties and filter out undesired ones—based on the explicitly defined features of the given F-set, and then to refine the filter by invoking implicative relationships, thus moving toward a more “theorem-based” algorithm. The limitations of the “filtering” approach stem primarily from the fact that, as mentioned earlier, the feature definitions recognize properties which are not necessarily comparable or commensurable with one another. In fact, only five of the eight features—namely G, DT, DP, BZ, and rounded or semi-reduced MP (where every generic interval appears in two consecutive sizes)—have compatible working definitions expressed in terms of restrictions on the G-set parameters c, d, and g. From the opposite perspective, this means that we are able neither to specify nor to eliminate from consideration the properties WF, ME, DE, or generic (non-rounded) MP. The ramifications of this quickly become apparent when one attempts to differentiate between scales belonging to F-sets 9 or 11, or to F-sets 10 or 12. Nevertheless, it is possible to construct meaningful and reasonably compact algorithms for all 12 feature sets.

Fortunately, the five commensurably defined properties are situated, for the most part, near the top of the implicative network shown in Example 2. As a result, the remaining “undefined” features can generally be accounted for (positively or negatively) on the basis of implicative relationships. Consider, for example, the case of F-set 1, which we have asserted includes only the usual diatonic. Given that any scale which is both BZ and DT will necessarily possess the remaining six properties, our assertion is easily shown to be true on the basis of the intersection of these two features.

For a more general model of the filtering-type algorithm and of the role played by implicative relationships among features, consider F-set 8, which applies to the generic triad or 7th-chord as embedded in the diatonic scale, discussed by Agmon and by Clough and Douthett. F-set 8 embraces all and only the sets with Agmon’s ELT property. The non-terminating algorithm for this family of pcsets is given in Example 4a. For simplicity, we omit here and elsewhere the additional steps required to recover all transpositions—that is,
all pcsets with identical values for c, d, and g. Given \( c = 7 \), then, the F-set 8 algorithm does generate the triad and 7th chord. However, it is by no means obvious that the algorithm produces either exclusively those pcsets which map onto F-set 8 or an exhaustive listing of F-set 8’s constituent members. That it does so becomes apparent if we examine the initial, rather crude “filtering” form of the present algorithm (see Example 4b) and then trace the refinements which lead to the present version. Because of space limitations, we omit discussion of this process here.

**Distributionally even sets.** Before concluding, we offer further remarks on the category we have called “distributionally even” beginning with an account of some of its more visible antecedents. These extend back at least to Messiaen’s 1944 compositional treatise with its discussion of his “modes of limited transposition” and range forward to Howard Hanson’s 1961 study devoted in small part to pcsets based on dual interval cycles. More recently, a line may be traced, more or less directly from the 1977-78 work of Starr and Morris, through Morris’s 1987 treatise, through Cohn’s 1991 paper, to our DE category. Along the way, one finds the 1982 paper of Lewin, whose generalized Riemann systems are necessarily conceivable as dual interval cycles, possibly incomplete, and the recently published characterization scheme of Anatol Vieru, based on “knots” of incomplete interval cycles. Two other contributions relating to the DE property are Stephen Soderberg’s study of Z-sets and Jay Rahn’s paper in which DE sets are studied in the rhythmic domain. In resonance with Richmond Browne’s 1981 discussion of the unique interval contexts for each pc in the usual diatonic, Rahn characterizes pcsets such as the octatonic scale as sets where one may “look out” from more than one pc (but not all) and see the same “panorama.” This metaphor perfectly describes the subcategory of our DE scales where at least one generic interval comes in one size and at least one comes in two sizes.

So there is an ample body of literature addressing Messiaen’s modes, transpositionally redundant sets in general, and the broad notion of the dual or multiple interval cycle. Our contribution along these lines is to define the category “distributionally even” in terms of generic interval sizes, and to open the category to cases where specific interval sizes are not rational parts of the octave. In so doing, we define a category that (and this came as a surprise to us) subsumes all the others in our classification except the deep scales and the generated scales with no other features. Like the category maximally even, it catches the pentatonic, whole-tone, diatonic, and octatonic sets; in addition, it includes all conjunctions of two complete interval cycles based on the same generator and many of more than two such cycles, thus embracing most of Messiaen’s modes and, more significantly, tunings of all the above that hold constant the distribution of interval sizes, for example, the Pythagorean diatonic.

The great generality afforded by the category DE leads us to suspect that further attempts to define categories in terms of generic and specific intervals (perhaps involving cases with more than two specific sizes for some generic sizes) may yield interesting results, but we leave that for the future.

**Conclusion.** Our principal objective has been to unify the contributions of many theorists who, motivated by a sense of wonderment at the diatonic scale, its durability, adaptability, and ubiquity, have taken account of its structure, and have gone on to pose questions extending to various musical dimensions, diverse musics, and the nature of musical systems, both extant and imaginary. We hope that the present effort will encourage and facilitate their continuing quest.
Example 1. Features of scales/sets

**features defined for rational or irrational generator**

- **G-generated**
  - May be generated by a single interval.
  - \( G(c, d, g) = \{0, g, 2g, ..., (d - 1)g\} \), products reduced mod \( c \), where \( c \) is the size of the modular unit (8ve), \( d \) is the cardinality of the set, and \( g \) is the generator.

- **WF-well-formed**
  - G-set in which each interval \( g \) spans a constant number of scalar notes.
  - \( G(c, d, g) = \{0, g, 2g, ..., (d - 1)g\} \), products reduced mod \( c \), where \( c \) is the size of the modular unit (8ve), \( d \) is the cardinality of the set, and \( g \) is the generator.

- **MP-Myhill-property**
  - Each generic interval (2d, 3d, etc.) comes in two specific sizes.
  - Each generic interval comes in either one or two specific sizes.

- **DE-distributionally even**
  - Each generic interval comes in either one or two specific sizes.

**features defined for rational generator only**

- **ME-maximally**
  - Each generic interval comes in either one size or two consecutive integer sizes.

- **DT-diatonic**
  - ME set with \( c = 2(d - 1) \) and \( c = 0, \mod 4 \).

- **DP-deep**
  - Every ic has unique multiplicity.

- **BZ-Balzano**
  - ME set with \( c = k(k + 1) \) and \( d = g = 2k + 1, k \geq 3 \)

Example 2. Implicative relationships

Example 3. Feature sets

<table>
<thead>
<tr>
<th>F-set no.</th>
<th>C</th>
<th>MP</th>
<th>WF</th>
<th>ME</th>
<th>DT</th>
<th>DP</th>
<th>BZ</th>
<th>example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>(0, 2, 4, 5, 7, 9, 11)</td>
</tr>
<tr>
<td>2</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>(0, 2, 3, 5, 7)</td>
</tr>
<tr>
<td>3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>(0, 2, 5, 7, 9, 11, 14, 16, 18)</td>
</tr>
<tr>
<td>4</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>(0, 2, 4, 6, 10)</td>
</tr>
<tr>
<td>5</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>(0, 1, 2, 3, 4, 5, 6)</td>
</tr>
<tr>
<td>6</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>(0, 1, 4, 5, 8, 9)</td>
</tr>
<tr>
<td>7</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>(0, 1, 3, 4, 6, 7, 9, 10)</td>
</tr>
<tr>
<td>8</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>(0, 2, 4)</td>
</tr>
<tr>
<td>9</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>(0, 2, 3, 6)</td>
</tr>
<tr>
<td>10</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>(0, 3, 5, 10)</td>
</tr>
<tr>
<td>11</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>(0, 1, 4, 5, 8)</td>
</tr>
<tr>
<td>12</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>(0, 5, 6, 11, 12)</td>
</tr>
</tbody>
</table>

Example 4. Algorithm for F-set 8

For \( G(c, d, g) \) as defined:

1. Set \( c = 3 \) (the minimum case).
2. Set \( g = 2 \).
   - Generate \( \{0g, 1g, 2g, ..., c - 3\} \) \( [d = (c - 1) / 2 \) in this case]\.
   - Generate \( \{0g, 1g, 2g, ..., c - 1\} \) \( [d = (c + 1) / 2 \) in this case]\.
3. Set \( c = c + 2 \). Go back to step 2.
b. Algorithm for F-set 8 (earlier version)
1. Set $c = 3$ (the minimum case).
2. Choose $d$ such that:
   - For $c$ even, $d = c/2$ or $(c/2) + 1$
   - For $c$ odd, $d = (c - 1)/2$ or $((c - 1)/2) + 1$
3. For each pair $(c, d)$,
   - Find $c'$ such that $cc' = -1 (mod d)$.
   - If some such value of $c'$ exists, find $d'$ such that $d' = (cc' + 1)/d$.
   - Set $g = d'$.
   - If $c'$ does not exist, discard the pair $(c, d)$.
4. Discard all $(c, d, g)$ where $c \equiv 0 (mod 4)$ and $c = 2(d - 1)$.
5. Discard all $(c, d, g)$ where $c = k(k + 1)$ and $d = g = 2k + 1$.
6. Generate $\{0g, 1g, 2g, ..., (d - 1)g\}$ (all products reduced $mod c$).
7. Set $c = c + 1$. Go back to step 2.

Bibliography


Wilson, Erv. 1975. Letter to John Chalmers re "the basic structure of moments of symmetry."

Inspired by educational reformer Pestalozzi, and communicating closely with such scientists as Helmholtz and Ellis, John Curwen developed in the mid-1800s a theory of "mental effects" and a set of hand signs to help teach sight-singing, music dictation, and tonal harmony according to his tonic sol-fa method. Arguably the most comprehensive approach to music training since medieval solmization, Curwen's pedagogy helped spread European principles to millions of teachers and students on every continent. It remains influential to the present through such adaptors as Kodály.

**Mental Effects**

Central to Curwen's method is the notion of "mental effects." For Curwen, "physical effects" comprise simple causal sequences, e.g., from physical frequency to phenomenal, perceived pitch. By contrast, "mental effects" involve "relations," e.g., among the consequences of such perceptual acts, as in the widespread musicians' distinction between relative and absolute pitch. Depending on its location within the major scale's interval pattern, each of the 7 syllables in Curwen's movable-doh system (doh, ray, me, ...) embodies a distinct mental effect or perceptual category: "doh-ness" or "ray-ness," as it were, as distinguished from a particular pitch.

That there are 7 such distinct categories in a diatonic scale is expressed by the phonetic uniqueness of each syllable, especially its initial consonant (d, r, m, ...) and can be justified in current theory by the fact that each degree occupies a distinctive place in the pattern, forming a unique interval configuration with the others. E.g., doh is 2, 4, 5 semitones below, and 1, 3, 5 above, the scale's other notes, whereas ray is 2, 3, 5 below and 2, 3, 5 above the others. In more recent terms, each degree has a unique "constellation" or "panorama." As in the medieval concept of modal affinities, the doh and ray constellations are, severally, more similar to soh's (2, 4, 5 below; 2, 3, 5 above) than to each other, a fact readily predicted by distances among these degrees on the "circle of fifths," or in current theory, in a cycle of transpositions up or down 5 semitones (T±5 mod-12).

Among the 7 syllables of Curwen's relatively modern system, doh is a p 5th below the phonetically similar soh, as is me below te. These phonetically parallel syllable-pairs form major 3rds, as does the phonetically similar pair fah-lah, inherited from medieval solmization and comprising the only other major 3rd in a diatonic scale. That there should be precisely three intervals of four semitones in the diatonic scale is also readily predicted by current theory.

The phonetically similar syllables me and te also draw attention to the fact that there are precisely 2 semitones in the diatonic scale, of which they form the lower components, "half an octave apart." This aspect of the diatonic scale is definitive in current formulations of "maximal even-ness" by Clough and Douthett. Moreover, the recent music-theoretical idea of maximal even-ness emphasizes ways in which such pitch-class collections as the diatonic, "white-key" scale are extremely privileged, i.e., in similarity/simplicity, by contrast with other possible interval patterns.

Curwen himself stressed precise interval patterns through his "modulator," an adaptation of Glover's "Norwich sol-fa ladder," a chart with syllables arranged vertically at distances proportional to the intervals between them. Curwen's graphic display was further correlated to (General) Thompson's fine distinctions in just intonation. Each syllable was associated also with particular words to convey as precisely as possible its mental effect within the entire scale. Especially in the earliest stages, students did not learn to read music in standard staff-notation. Instead, Curwen devised a "New Notation," employing the first letter of each syllable plus diacritics and punctuation (apostrophe, colon, period, etc.) to denote precise values of register and rhythm. His aim in formal music training was to develop and exploit as far as possible mental effects, i.e., the "ear" of the neophyte musician, as a basis for learning.
Handsigns Curwen also introduced handsigns to convey relative pitches and their mental effects without recourse to graphic, typographic, spoken, or sung syllables. Especially valuable for conveying simultaneous tones in a classroom, the individual handsigns were uniquely associated with particular mental effects, scale degrees, syllables, etc. Adapted to Kodály’s 20th-century curriculum, such manual signs are still used widely in Europe and North America.

Plausibly because of great speech divergences among his students and his considerable reliance on printed materials, Curwen’s textbooks emphasized clear distinctions in pronunciation. Taking as his point of departure Ellis’s work in phonetics, Curwen explicitly conceived syllables as articulative, not merely sonic, and implicitly anticipated “motor phonetics” and recent work in the computer simulation of speech and song. In this way, Curwen’s handsigns can be understood as a relatively “large-motor” approach to grasping mental effects, additional to the relatively “fine-motor” resources of syllables.

Triadic Tonality and Curwen’s Handsigns I believe Curwen’s handsigns comprise structures that illuminate one of the most difficult distinctions in music theory, namely, between scale and tonality. Drawing largely on the mathematical resources and findings of atonal music theory, recent accounts of scales, chords, and intervals have not yet responded fully to certain questions raised—or at least, begged, but certainly never answered adequately—by traditional tonal theories. E.g., in general—for generality is theory’s defining specialty—how can one distinguish among, and inter-relate, consonance, dissonance, and resolution?

An outstanding, somewhat recent exception is Boretz’s informal “outline of a musical system,” already of great importance in the intellectual history of music because of its aim to connect tonal and atonal/serial theory. Although Boretz seems not to have been immediately aware of either as a precursor to his own system, pitch-class structures he delineates are clearly isomorphic with motor patterns among Curwen’s handsigns and with standard discant/counterpoint techniques formulated and carried out as “common practice” by theorists, composers, performers, and improvisers of “modal” (arguably, “dyadic”) music in the Middle Ages and Renaissance. In what follows, I attempt to extend in generality Boretz’s formulation of triadic tonality. I draw into this extended discussion other, more recent studies in music theory. I also try to determine what aspects of music might be learned, albeit non-verbally, through Curwen’s handsigns.

The main dichotomy in Curwen’s handsigns is between the tonic triad and the other 4 degrees of the major scale. For tonic-triad degrees, the hand is continuous with the forearm, wrist unbent: for m, the palm is extended in the horizontal plane; for s, the same, but vertically; for d, a fist in m’s horizontal plane. Within this group of 3 handsigns, m is medial in similarity between d and s, which contrast in the oppositions fist/palm, and horizontal/vertical.

These contrasts can be understood not just visually and “objectively” (e.g., from a spectator’s viewpoint) nor only statically, but also kinesically, dynamically, and (inter-)subjectively, especially in transitions from one to another. Successive pairs of handsigns involve more or less similar directions/orientations and changes of particular muscles and tendons, e.g., the forearm’s twist between s and m or d. The 4 non-tonic-triad degrees form melodic dissonances with the tonic triad and engage further motor contrasts. In all four, the wrist is bent or the forearm is twisted, and one or all fingers point to a single resolution (respectively, d, m, m, or s). The index finger points upward for t, downward for f, whereas all digits point upward for r, downward for l. Cross-cutting the upward-downward dichotomy, is a semitone/whole-tone dichotomy between the index finger and all fingers (‘<‘ vs ‘>=‘): t->d, r=m<s<, s<s<. Finally, both t and l require a twist of the forearm in proceeding to or from their resolutions (d and s, respectively).

For Curwen, major tonality is more important than minor or other “modal” structures. Pre-eminent in any key are the triads on degrees d, s, and f (for which Curwen’s chord symbols are D, S, and F). Accordingly, Curwen’s handsign pattern can be understood as follows: t->d, r=>m, s<, m=-f.
s<=l; i.e., each degree in $S$ (t, r, s) and $F$ (d, f, l) resolves stepwise to a particular degree in $D$ (d, m, s), except $s$ in $S$ and $d$ in $F$, which, as degrees in $D$ itself, are already resolved. In more recent terms, whereas the octave modulus orders particular degrees in a one-dimensional, low-high cycle throughout all registers (d r m f s t d r ...), Curwen's resolution pattern privileges 3 of these degree-classes (d, m, and s) and groups each of the other 4 degree-classes with one of these in an ordering, extending across all registers but bounded/broken by the opposite directions of $t$ and $l$.

In general, an individual melodic dissonance is grouped with, and its handsign points toward, the tonic-triad degree-class to which it is closest. It might seem arbitrary to group $r$ with $m$, rather than with $d$, for in current theory both are simply 2 semitones away from $r$. However, within just intonation, which Curwen adopted as his tuning framework, $r$ ideally forms a smaller ratio with $m$ ($10/9=81/72$) than with $d$ ($9/8=81/72$). More important, I believe, grouping $r$ only with $m$ preserves the determinacy Curwen sought in general (plausibly for pedagogic ease). Even more important, I feel, this grouping fulfills a symmetry implicit in Curwen's triadic ideal, irrespective of the nice tuning distinctions he sought to inculcate (albeit as a secondary goal of his curriculum). In any event, whether understood in Pythagorean, just, or equally tempered terms, Curwen's resolution pattern rings true with the tonal practices he aimed to codify and teach to neophytes.

In semitone-sizes, dms, fld, and str are identical: 047. Hence, each is a transposition of the others: fld is $T3dms$; str is $T-5dms$. In generic, scale-degree intervals, these triads are also identical (024) and transpositions of each other: fld is $T3dms$; str is $T-3dms$. More precisely, dms, transposed up 5 semitones/3 scale degrees and down 5 semitones/3 scale degrees, yields all and only the notes of the diatonic collection: drmfslt. Curwen explicitly specifies this relation between the three triads and the entire scale as a basis for according these central sonorities privileged status within his theory of harmony.

Whereas one could list many isolated features of Curwen's triadic pattern of melodic resolution, a more unified account of their inter-relationships results from describing systematically the general pattern of which the handsigns are an instance, and conversely, by identifying traits that distinguish it from other, seemingly cognate possibilities. In this way, one can begin to specify more precisely the "class" or "category" of motor routines learned in using handsigns.

**Mathematical Aspects**

Musically, Curwen's pattern can be understood as embodying voice-leading connections, i.e., a cross-registral "trio" of degree-class "voices." Other plausible structures include a "solo," "duet," "quartet," "quintet," ..., comprising a similarly centric, interleaved arrangement of monads, dyads, tetrads, pentads, ..., i.e., of "chords" having 1,2,4,5, ... degrees. Mathematically, if the number of degrees in the central, "tonic" chord is $n=(1),2,3,4,...$, the number of degrees in the scale it generates by transposition must be $3n-2$ (since both non-tonic chords share a unique common-tone with the tonic chord). Registrally consecutive tones in each chord can be no more than 3, and no fewer than 2, degrees apart (because of constraints comprising adjacency and registrally directional mapping of the central chord one-to-one and onto the other two). The number of 2-degree intervals in a chord must be 2 (for if this number is set at $p$ and the number of 3-degree intervals is $q$, $2p+3q=3n-2$; however, $n=p+q$; thus, $p=2$ and $q=n-2$). In general, chords having more tones have more intervals of 3 degrees. Most precisely, Curwen's pattern requires that the central, "tonic" chord consist of an odd number of tones (i.e., $4n+1=(1),5,9,13,...$ degrees, this "medial" degree is not registrally "on the way up from" the chord's "root" to its "fifth," as its major half-octave comprises an odd number of consecutive intervals, specifically, $((4n+1)+1)/2=2n+1$.) In the well-known instance of diatonic, 7
scale degrees are distributed maximally evenly among $2(7-1) = 12$ semitones (specifically, for major: 2212221 semitones upward). This, and the corresponding match of semitone-sizes among the diatonic scale’s three central triads—whether major (047) or minor (037) —suggests, especially for chord-scales of $4n+1=7,11,15, ...$ degrees, the possibility of distinguishing, more generally, between major and minor along such lines. At the level of scale-degrees (in contrast to semitones, or more generally, “microtones”), there is no difference between the “thirds” within such chords. Nonetheless, each of these scalar chords contains not only such “neutral thirds” as 5 (=2+3 and 3+2 in the 7-degree chords of the 19-degree scale), but also such contrasting versions as 6 (=3+3).

Indeed, because they comprise three minimally overlapping, maximally even chord-scales, such structures can “fractionate” into three subscales, each internally manifesting the Curwen resolution pattern (complete with the major-minor distinction, e.g., between 6 and 5, as well as relatively narrow “t-d” and “m-f” intervals etc.), but centred on a different tonic-chord (e.g., a chord-degree higher) than the larger scale they comprise:

```
1 --> 3 6 --> 9 --> 11 14 --> 17 [1 --> 3 6 --> 9 --> 11] 10 --> 12 15 --> 18 --> 1 4 --> 7 [10 --> 12]
```

This structural disjunction corresponds to the requirement for minimal overlap, which distinguishes transpositions that yield Curwen’s resolution pattern within a scale from those that yield chromaticism or modulation beyond a scale, where, for example, overlap cannot sink to such a minimum for diatonic construed mod-12). As in Agmon's formulation of chord-relations, adjacent degrees and common tones play a basic role. In contrast to Cohn’s formulation of chords that cycle infinitely stepwise within a scale, Curwen structures implode and explode “fractally,” as it were: hence the inclusion of 1-degree “chords,” comprising 0-degree intervals—not just a mathematical curiosity, but a structural limit constituting “closure” in a totally ordered hierarchy.

**Empirical Hypotheses** How might one conclude that a Curwen pattern is actually functioning, e.g., in an experimental setting? Such a structure is syntactical and hence no more immediately observable than the grammar of a language or dialect. Moreover, just as in the production or perception of speech, where more than one syntax can compete, other musical systems (e.g., dyadic/modal) can vie with triadic tonality of the Curwen type. Nonetheless, in analyzing (e.g., probe-tone) data, one would expect a central chord to dominate responses if Curwen tonality is effective. Further, one would also expect neighbour-tone responses, not merely as parts of a central chord’s complement within a scale, but also as unresolved portions of the central chord’s voices. Indeed, if chords and scales functioned independently as agents of probe-tone responses, such alternatives as fah and fe, and te and tay would be equiprobable as responses to such “gapped,” elliptical stimulus-sets as drmsl. By contrast, all other factors held constant, fah and te are to be expected if Curwen symmetry is a reinforcer.

If similarity/simplicity relations are reinforcers, common chords should be relatively favoured over diminished triads. 3 major vs 3 minor triads should be favoured behaviourally as subgroups over other divisions of these 6 into 2 groups of 3; similarly for other “maximally frequent” chords in other, microtonal systems, for which Clough and Myerson precisely predict multiplicities. In this sense, triadic voice-leading is a capacity of diatonic “waiting to be reinforced.” As Boretz’s account suggests, “trio” frameworks, as such, actually favour such sonorities as stf and (it inverse) tfl over trf, and even over rfl in major and str in (natural) minor. Whereas delicate non-pitch (i.e., timbral, micro-rhythmic/loudness) factors are all one can cite in defence of Curwen’s apparently Euro-centric privileging of major (as well as root position, dominant skewing, and raised leading-tones), one must acknowledge that Curwen’s formulation itself has been for more than a century a worldwide factor in shaping production, perception, cognition, and conceptualization at all levels of formal training. Arguably, the closed, “totalizing,” hierarchical structure of Curwen tonality has been a powerful agent of European aesthetic hegemony, enabling quick “overlearning” of authoritative, tonic structures, while erecting barriers to subsequent fluency in “other” possibilities.
REFERENCES

Agmon, Eytan (1990) Linear Transformations between Cyclically Generated Chords Musikometrika 3: 1-26


Clough, John; Gerald Myerson (1985) Variety and Multiplicity in Diatonic Systems Journal of Music Theory 29/2: 249-70


Cringan, Alexander T. (1889) Teacher's Handbook of the Tonic Sol-Fa System Toronto: Canada Publishing

°(1861) How to Observe Harmony 14th ed. London: J. Curwen & Sons (1913)

Ellis, Alexander J. (1850) Speech in Song London: Novello, Ewer and Co's Music Primers

Glover, Sarah (1835) Scheme for Rendering Psalmody Congregational Norwich, UK: Jarrold


°(1983) A Theory for All Music: Problems and Solutions in the Analysis of non-Western Forms Toronto: University of Toronto Press
°(1991) Coordination of Interval Sizes in Seven-Tone Collections Journal of Music Theory 35/1-2: 33-60


Curwen patterns for 3-, 5-, and 7-degree "tonic" chords

n=3  \(3n-2=7\)  chord-cycle: 2 2 3

resolutions: 0  \(\rightarrow\) 1 \(\rightarrow\) 2 \(\rightarrow\) 3 \(\rightarrow\) 4 \(\rightarrow\) 5 \(\rightarrow\) 6

"dominant" T+4  0  2  5

"tonic"  1  3  2

"subdominant" T-4  4  6

n=5  \(3n-2=13\)  chord-cycle: 2 3 2 3 3

resolutions: 0  \(\rightarrow\) 1 \(\rightarrow\) 2 \(\rightarrow\) 3 \(\rightarrow\) 4 \(\rightarrow\) 5 \(\rightarrow\) 6 \(\rightarrow\) 7 \(\rightarrow\) 8 \(\rightarrow\) 9 \(\rightarrow\) 10 \(\rightarrow\) 11 \(\rightarrow\) 12

"dominant" T+7  0  2  5  8  10

"tonic"  1  3  6  9  11

"subdominant" T-7  4  7  12

n=7  \(3n-2=19\)  chord-cycle: 2 3 3 2 3 3 3

resolutions (transpositions T±10):

0  \(\rightarrow\) 1  2  \(\rightarrow\) 3 \(\rightarrow\) 4  5 \(\rightarrow\) 6 \(\rightarrow\) 7  8 \(\rightarrow\) 9 \(\rightarrow\) 10 \(\rightarrow\) 11 \(\rightarrow\) 12  13 \(\rightarrow\) 14 \(\rightarrow\) 15  16 \(\rightarrow\) 17 \(\rightarrow\) 18

Curwen handsigns and mental effects for the diatonic scale (after Cringan)

SOH: The grand or bright tone

TE: The piercing or sensitive tone

ME: The steady or calm tone

RAY: The rousing or hopeful tone

LAH: The sad or weeping tone

FAH: The desolate or awe-inspiring tone

DOH: The strong or firm tone
In order to test a basic assumption in tonal analysis, a Schubert waltz — a very short piano piece with quite modest melodic and harmonic excursions — was manipulated so as to end in another key than that in which it started. The interferences as such were as unobtrusive as possible in order not to offend the listeners' sense of good local continuation, but the effect was nevertheless devastating theoretically speaking: the controlling tonality was upset and so was tonal closure.

Schubert's waltz conforms to B minor: after a transient cadence to the dominant F# major in m. 4, it settles in the relative key of D major at the double bar in m. 8, from where it finds its way back to a full cadence in B minor. The original piece was recomposed so as to remain and close in D major; melodically the second part of the waltz was changed so as to return to F# before a final descent to D. Since this first variant is ambiguous in a way that might somewhat obliterate the mismatch between beginning and end — rather than hearing a wrong D-major conclusion of a B-minor piece, listeners might (if noticing anything at all) hear a D-major piece with an odd B-minor beginning — a further variant, giving more emphasis to the initial B-minor tonality by rounding off the first part of the piece in the tonic, was also prepared. This second variant might however also be understood in a way that smooths over the mismatch: it might suggest an ABA-structure in which the closing B-minor part is missing.

Schubert's composition as well as the two variants of it are given as Schenkerian graphs in exs. a, b, and c; above the graphs are shown prolongational reduction trees according to Lerdahl and Jackendoff (1983). The original waltz agrees excellently with Schenkerian principles and yields both an impeccable Ursatz and a "basic form". The underlying structures of the variants, however, do not qualify as Ursätze, whereas the corresponding trees may still describe "normative structures", though the dashed lines indicate the uncertainty of some connections.

Leaving theory for empirism, the variants were judged by five groups of conservatory/musicology students according to the following scheme:

- Group 1 listened to variant b with repeats, and were asked to write down their judgement after just one presentation
- Group 2 heard variant c under these same conditions
- Group 3 listened to variant c with repeats on five occasions distributed over nine days, but were asked to comment on the music only after the last presentation
- Group 4 heard variant b without repeats three times in one session; comments were collected after each presentation
- Group 5 listened to variant c under these same conditions

All subjects were given this instruction: "You will hear a short piano piece played in its entirety. Listen very carefully!" In addition their listening was guided in the following way. Before listening, the subjects of Groups 1-2 were given the clue that "something is odd with the harmonic/tonal development of the music — find out what it is!", whereas this clue was given to Group 3 only before the last presentation. As to Groups 4-5, the listeners were first just told to listen attentively, then that "there is something strange
and unusual with the piece”; only before the third presentation were they given the explicit clue.

When working through the results, all comments that could be taken as correctly catching the crucial deficiency of the music were noted. Subjects reporting that they recognized the piece or that they possessed absolute pitch, were excluded from the investigation. The outcome was as follows — chi²-calculations are used to estimate the statistical significance at probability levels .05* and .01**.

Group 1 (N=14; variant b) — 2 hits**
Group 2 (N=18; variant c) — 2 hits**
Group 3 (N=16; variant c) — 3 hits*
Group 4 (N=21; variant b) — 0 hits** then 5 hits* then 4 hits
Group 5 (N=19; variant c) — 0 hits** then 2 hits** then 4 hits

Despite the clues given, the listeners largely failed to notice the tonal mismatch between beginning and close and were not offended by the lack of tonal closure. This conclusion holds for both variants, though variant b might perhaps have been somewhat more disturbing than variant c. Rehearing as such did have little influence; only when the variants of the waltz were played without repeats, and when the presentations were combined with gradually more outspoken clues, did the ratio of relevant comments raise to a substantial level — and yet more than half of the subjects were not able to identify what was wrong in the music.

The listeners turned out not to demand a return to the "home key" when the variants approached their final cadences. It should be pointed out that the listeners in question were far from inexperienced, that they were incited to notify the tonal course of the music, and that some of them were given the advantage of listening more than once.

Turning to the original Schubert waltz itself, it is not overly long and exhibits a limited modulatory complexity. Generalizing somewhat — a short Mozart movement was also tested in the way just described and yielded a similar outcome — the result indicates that listeners' capacity to detect mismatches between beginning and concluding key is quite poor. This means, in other words, that the long term memory for the tonic is severely restricted. But to remember such a thing seems to be a difficult task, so isn't after all this negative result what could be expected?

The problem is that much music theory has made far-reaching commitments to the contrary. The fact (or rather the non fool-proof generalization) that all pieces of tonal music return to the tonic, has been taken as evidence for the assumption that return to the tonic is a basic and necessary aspect of global closure and tonal coherence, indeed a defining property of tonality itself. One conclusion to be drawn from the present experiment is therefore that the unifying effect of tonality has been exaggerated. The vast majority of tonal pieces do state a key to which they return, but this does not entail that they are primarily heard — or that they must necessarily be described — with this fact in mind. Tonality is just as much, or perhaps rather, a local phenomenon that makes itself felt in diverse ways, and musical unity is certainly brought about by many other factors besides tonal coherence.
More specifically, it is a central assumption in Schenkerian theory and a prerequisite for "classic" Schenkerian analysis that pieces of tonal music begin and close in the same key. The idea of a single key, a primordial chord, dominating all progressions in a music work and imparting unity to it, is regarded as the core and essence of tonality. And the existence of a persisting background key lends credibility to the Ursatz, that governs and informs the musical organism, and justifies the analytic endeavours to recover this fundamental structure from the depths of the musical design.

Apart from ambiguities due to the fact that the Schenkerian tradition is old enough to have a history of translations, appropriations, developments, and revisions, Schenkerianism seems to be both a theory about how (even-deficient tonal) music is actually constituted, and a theory about what (competent) music listening is or should be. The reductions (or perhaps rather the evolving prolongations) describe how Masterpieces are heard by Initiates, and listeners with exacting ears demand unity, demand that the musical surface is controlled by a hierarchy of contrapuntal/harmonic frameworks, all related back to the Ursatz.

Methodologically speaking, listening is assigned two complementary (actually to a considerable extent interdependent) functions in Schenkerian analysis: listening is used (or said to be used) when detecting the higher-level structures — a guide for the selection of events in the reduction process — and when verifying what has been found. The latter use as Satzprobe need not be considered at any length here since it seems to be an illusory procedure: is it really possible to listen to abstractions? Furthermore, to check whether the structures obtained make musical sense by conforming to the rules of strict counterpoint is more or less superfluous, because correct voice-leading ranks very high among the rules determining permissible reduction, and also unsufficient, since what should be checked aurally isn't, namely the relationship between the deep structure and the surface configurations in which it is claimed to reside. The former, probing use of listening, however, is doubly crucial since it involves not only the origins, and thus the experiential legitimation, of individual reductive readings, but the validation of the entire theory.

For one of the props supporting Schenkerian theory is the fact that a great number of compositions have been found to conform to their respective keys — when examined according to the principles of Schenkerian analysis, all these pieces turn out to embody just the kind of structures stipulated by that very theory. Since this seems impossibly circular, an independent source of evidence (a prop supporting the prop) is badly needed. And listening — mainly a beginning-to-end, bottom-up activity — is a suitable instance to advance in defense against the objections sometimes raised that Schenkerian analysis is based on visual inspection of the score (traits of which the analyst then, according to the principles of the theory, highlights or downgrades), and that it is too much a time-less, top-down procedure to do justice to an art that exists in and dynamically changes over time. What happens to these props, when it turns out that ordinary, good listeners loose the track, forget the tonal point of departure and accept final cadences that can't close the music according to the Schenkerian concept of tonality?

As far as the forming of a mental representation, however sub-conscious and however faintly resembling a Schenkerian deep middleground or background, is a prerequisite for a sense of tonal coherence — the hierarchy of reductions leading back to the Ursatz.
may be thought of as a machinery making memory of the tonic possible — most listeners are obviously not able to suspect, let alone arrive at, such a representation. To disqualify ordinary good listeners, to claim that true appreciation of tonal coherence, and consequently the ability to verify Schenkerian accounts of musical structure, is a precious gift that is only bestowed to the happy few, or that it takes years of specialized instruction to learn to listen in a way that matches the theory, is a problematic defence. It makes Schenkerian analysis a very exclusive activity and — what is worse — exposes the theory to the suspicion that the phenomenological support for the analyses cited as evidence for the theory in fact does make up a vicious circle: fundamental tonal structures can be aurally found and corroborated only by those who have accepted the theory and been trained in applying it. And it paves the way for a further suspicion: to what extent are the connections set out in Schenkerian deeper level graphs really heard at all — even by the analysts themselves?

At this point an excursion to related research is appropriate. Nicholas Cook (1987a) has undertaken an experiment, similar to the present one, involving listener reactions to pieces changed so as to close in the wrong key. His results are consonant with those reported here: listeners don't notice, and are not offended by, such tonal deteriorations. Vladimir Konecni and Mitchell Karno (1994), reviewing various experiments confronting cherished analytic truths with listeners' aesthetic verdicts, approvingly cite Cook's investigations. When it comes to the interpretation of these findings, however, their conclusions diverge.

Konecni and Karno argue for wholesale condemnation: behavioural science is given the prerogative to determine what is possible to assert and permissible to propagate in music theory. Cook's attitude is entirely different. "If the principle of tonal closure has little or no perceptual validity at the larger time scales found in most tonal compositions, is there not something radically wrong with a theory that ascribes fundamental aesthetic importance to it? To ask this is to assume that a theory of musical structure has to be also a theory of perception. But there is no intrinsic need for the theorist to conceive of musical structures in the same manner that the listener perceives them." And indeed, when teaching how to arrive at Schenkerian interpretations, Cook (1987b) assigns a modest role to listening — the Kopfnote is best found if you read the music backwards. In a later contribution (1989), discussing the scientific status of music theory, Cook pleads in favour of Schenkerian analysis. Reductive graphs are verified by each individual reader's consent, and the value of an analysis stems from the fact that it — to the extent that it does not reduplicate aspects of the surface — provides an illuminating "comparison" with the actual music.

Granting that analytic theories and observations should not — without qualifications and empirical corroboration — be presented as valid for listening in general, there is an element of over-kill in Konecni and Karno's position. On the other hand, it seems that Cook lets Schenkerian analysis off the hook too easily. It is salutary that listening in the emphatic, Herculean sense is demystified and removed from Schenkerian theory as a heuristic device — Cook manifestly completes the covert retreat that (the title notwithstanding) is felt already in Salzer's "Structural Hearing" — but the scope and nature of the concession must be assessed. The principle of tonal closure, the Ursatz, is not only considered to be aesthetically vital — Schenkerian theory necessarily ascribes fundamental methodological importance to it. If the existence of the Ursatz (along with much of the middleground activities prolonging it) is perceptually unimportant when
you listen, why should it govern the analyst's eye? Is the metaphysical foundation for it really strong enough to support the element of top-down processing in tonal analysis? And what else is left to recommend it when the collected analytic evidence from successfully reduced Masterpieces can be questioned?

Turning to Cook's pragmatic attitude to verification and evaluation, he is no doubt right when asserting that there is no intrinsic need for an analytic theory to model perception. But on the other hand we should be wary not to adopt the notion that there is no intrinsic relationship between the value of a theory and the extent to which it deals with the musical phenomena we hear, indeed with the musical surface we see? How much non-duplication can an analytical description take and still emerge as illuminating? Can an analysis really be enlightening if it doesn't fit? Is faithful reproduction of actual musical traits really that inimical to musical understanding? Don't top-down prejudices, programmatic disfavouring of certain design parameters, and disregard of surface salience sometimes entail a very high price? And we should not forget that there is also a hidden cost involved: the insights that might have been gained from other, less "tonal" reductions, from descriptions that are suppressed by analytic orthodoxy.

References:
Cook (1987a) "The Perception of Large-Scale Tonal Closure", Music Perception 5/2, 197—206
Composition and Improvisation
A typology of compositional styles in computer-based music making
Dr. Göran Folkestad, Malmö Academy of Music, Sweden
Dr. David J. Hargreaves, Leicester University, U.K.
Dr. Berner Lindström, Göteborg University, Sweden

Introduction
Recent technological developments and the increasing impact of the media mean that listening to
and creating music constitutes a major and integrated part of many young people’s lives.
One result of technology is that attitudes towards the creation of music have changed, and are
more egalitarian. Creating music is no longer seen as reserved for geniuses, but as an activity in
which everyone can participate. Using an artefact such as the computer might be seen as an
embodiment of this changed conception of music.

Kratus (1985) investigated the developmental nature of children’s original compositions. Eighty
children aged 5 to 13 composed melodies on hand-held electronic keyboards. In the analysis,
Kratus focused on rhythmic and melodic patterns, and phrase characteristics of the music. In
addition to his work on creative products, Kratus (1989) carried out an investigation in which
the compositional processes used by children of ages 7 to 11 were analysed. A small electronic
keyboard was used, and the children were restricted to use ‘only the white keys’ and to have
middle C as the starting pitch. Kratus recorded not only the resulting composition, but also the
10-minute experimentation that led to it. The findings ‘suggest developmental differences in
children’s strategies for composing music’ (p. 17).

Polkestad (1991) carried out a two year study in a secondary school in which 12 year old pupils
working in groups of 3-5 created music of their own, using synthesisers and sequencers. The
results showed that the implementation of music technology not only changed what was done,
with respect to musical content and styles, but also how it was done, and two qualitatively
different ways of creating music were identified. These were defined as (a) supplementary use,
in which the equipment was used as a tool for arranging the music, and (b) integral use, in
which the equipment was used as an interactive medium from the very beginning as an integral
part in composition.

The present study: Methodological issues
The design of the present study aimed at creating a more open-ended situation than in Kratus
(1985, 1989), a situation similar to the context of the out-of-school activity, in which young
people create music of their own. By leaving out the teacher and the educational context as
much as possible, the aim was to investigate the creation of music as it appears in informal
learning situations outside school. The aim of this paper is to describe the composition process,
by presenting a typology of compositional styles in computer-based music making.

observation of composers during a session of composition’ (p. 103). As he points out, ‘this
requires a rare degree of co-operation from a composer’ (p. 103), which has made this an
uncommon method in practice. Another problem with observations is that the presence of the
observer changes the situation under investigation. That problem is exacerbated when studying
novices who have not yet found their identity as creators, and are thus more easily affected by
the presence of another person, especially if that person is regarded as an expert in the field.

In this perspective, music technology was found to be helpful, as a research tool in trying to
achieve the aims sketched by Sloboda, namely to get access to ‘the notes written down, their
sequence and grouping in time’ (p. 103), without being constantly present as an observer. The
option used was to continuously save the MIDI information from the creating process in the
computer files, and thereby get the chance to follow, analyse and compare the course of events
afterwards, from the first idea until the completed piece of music. One important advantage of
this computerised portfolio-method (Wolf, 1988; Gardner & Perkins, 1989) is that the
participants can work in privacy, without having another person watching them in close
proximity.
Method
The present study was part of a three year project, in which a total of 14 participants, eight boys and six girls, aged 15-16 and with no previous experience of composition, worked individually after school once a week, with no teacher instructing them.

The technique was to use the save as-command, instead of the save-command, which obliterates the stage before. In practice that was done by changing the name of the document each time a new event was about to be saved, simply by changing the letter behind the title of the song.

In addition to this, interviews were regularly carried out each time a composition was completed, in which the participants described how they had worked and what thoughts underpinned their actions. The basic focus in these interviews was consequently on their activity, and their thoughts about it. The interviews were carried out at the work station. At the start of the interview the researcher and the participant listened to the composition together, and the participant was also presented with earlier versions of the same composition to comment upon during the interview.

The work stations consisted of one computer with a sequencer program (Master Tracks Pro by Williams & Howell, 1990), one multi-timbral synthesiser, and headphones. The demonstration of the equipment, lasting for about 15 minutes, included how to use the sequencer program as a multi-track recorder, how to choose sounds from the multi-timbral synthesiser, and address midi channels. All other instructions about the options in the program were made later during the work, on the initiative of the participants, and when asked for in their questions and problems. No manual of the program or any written instructions were given.

A total of 129 pieces of music were created, and during that work 887 MIDI-documents were saved. For each piece of music, the saved documents were listened through in order to sort out the process of creation. The recorded interviews were transcribed verbatim, as well as the notes from the observations.

In order to assess the inter-judge reliability of the categories of composition, two judges were selected, and a sample of 15 compositions were randomly selected, making sure by allocation by quotas that all the categories were represented in the sample. The computer files for each of the selected compositions were coded with numbers, thereby taking away any information that could reveal whether the composition was made by a boy or a girl, or if any of the compositions were done by the same participant. The parts of the interviews dealing with each of the compositions were selected and encoded as described for the computer files. The results of the test were adequate in both cases; the percentage agreement between judge 1 and the researcher was 67%, and between judge 2 and the researcher 73%.

Results: The typology of compositional styles
The following is a summary of the main account of the typology of composition strategies already published in separate articles (Folkestad, Hargreaves, & Lindström, 1997) and Folkestad (1996). The main account describes in full how the categories were developed, and also deals with issues of validity and reliability.

The choice of strategy, or way of working, HORIZONTAL or VERTICAL, denotes the main difference between the modes of composition.

Horizontal composition
As a first phase in horizontal composition, the song is completed in form and content from beginning to end. After this, the computer equipment is used for arrangement and instrumentation of the composition. The strategy, as it works conceptually, and also as it is visualised on the screen, implies that horizontal lines are being completed one by one, from beginning to end. According to this, a basic feature of the horizontal categories is that composition and arranging are separate processes.
Horizontal 1a: Composing at an instrument - arranging in front of the computer. This category is distinguished by the composition being produced as a unit by playing, listening and evaluating. This integrated process is not interrupted until the composition is completed as a whole. Subsequently, the composition is recorded on one track on the computer. Thus, the process represents a totality that is not interrupted until the entire composition, including melody, chords, and form, has come to an end.

This way of creating music presupposes that the composer’s instrumental skill in playing the keyboard is good enough to be able to realise the musical ideas. Creating the music in this way also means that the piece is learned by heart, and rehearsed so well that the result of the recording seldom has to be corrected.

Horizontal 1b: Composing at an acoustic instrument using the computer as co-musicians. This category is characterised by the use of an acoustic instrument as a complement to the computer equipment, both in the actual composing, and in arranging/orchestrating.

In the first phase of work, the composition is completed on the acoustic instrument. However, the finished composition is not transferred to the computer via the keys of the synthesiser, as was the case in HORIZONTAL 1A. The original composition is thus always to be found ‘outside’ the computer, and the different harmonies in the arrangement are tried out one by one on the instrument, and then transferred to the computer via the keyboard.

The equipment thus functions as co-musicians in a band when a preliminary musical idea is tried out at the rehearsal, testing and developing it with the help of the various instruments in the band. The arranging is done by testing and recording track by track, and in this way the tune is built up continuously by playing ‘live’ on the acoustic instrument along with the computer instruments, and track by track recording how ‘the others’ are to play.

Horizontal 2: Horizontal composing piece by piece in front of the computer. As in the other categories of horizontal composition, work starts with the whole composition being conceived to completion; in this case the computer equipment is utilised from the very beginning. Composition is done by trying different things out, piece by piece, until the melody and/or harmonies of the composition are correct from beginning to end. This strategy, both conceptually and also as visualised on the screen, implies that the first horizontal line representing the actual composition develops bit by bit. These bits or pieces do not have to be musically organic parts, like verse or refrain, but are merely pieces of varying length recorded and interrupted at any point, and put together as a whole during the creation of the tune. As in HORIZONTAL 1, the subsequent instruments in its arrangement are then completed one by one, from beginning to end.

Accordingly, the main difference between HORIZONTAL 1 and HORIZONTAL 2 is in the way the form and contents of the composition are developed. In HORIZONTAL 2 this is done by playing/recording some bars with a varying degree of improvisation and deliberate testing, and then letting the computer equipment replay the recorded part; one listens, evaluates and retains what is good before continuing in the same manner with the next part.

Vertical composition
In vertical composition, the various sections of the composition are completed with all instruments before moving on to the next. When the first section is completed, there is as yet no idea of what is to follow, nor of the number of sections in the final composition. This is defined while composing. Thus, both composition and strategy take the form of vertical ‘chunks’, in which each chunk, comprising an organic musical part, is completed for all instruments before moving on to the next phase. The vertical way of thinking is shown by the way instrument structure is defined at an early stage for each phase (VERTICAL 1), or from the very start regarding the entire composition (VERTICAL 2). As a conclusion of this, composition and arrangement/instrumentation are conceived as one integrated process within the vertical strategies.
In this work, vertical thinking finds its expression not only in working (moving) from top to bottom, but also by frequently back-tracking to re-record previously recorded instruments, this time with the more recent tracks as a reference. In this way, thinking — and the piece itself, accordingly — moves up and down between the various instruments, forming a vertical unit in the section being composed. This work is based on improvisation, and the possibilities provided by the equipment form the basis of this method of musical creation.

Thus, the computer is used as an interactive medium to a greater extent than that described in HORIZONTAL 2, in that the sounding response from the equipment to an idea being tried out predominantly guides the development of the subsequent composition. The distinction could be described as follows: In HORIZONTAL 2, the task of the equipment was to represent the recording precisely in the way it was recorded — like a tape recorder; while in Vertical composition the possibilities of the equipment are also utilised in processing the recorded material before it is replayed over again. In this way, the equipment is not merely used as an aid in the composition work, but is integral to this way of working.

**Vertical 1a: Vertical composition section by section.** This category is characterised by the formation of various parts, which are clearly discernible musically. When some of these parts are completed as described, the copy-paste function is utilised in order to give the composition its final form. The instrumentation emerges part by part during the composing of the various parts. These are often made up of different ‘orchestras’, each playing its own part, and contrasting with each other, sometimes in an almost concerto grosso-like manner. All these parts are held together by a drum accompaniment, which in most cases is the same all through the piece, and consequently turns out to be what holds the piece together stylistically.

**Vertical 1b: Vertical Composition as Sound Composition.** In this category, the sounds and the created sound structures have a central function, and the musical pieces being created could chiefly be described as sound compositions in which the concrete sounds offered by the synthesiser are utilised. Because the tonal picture is essential in the creative process from the outset, and actually provides the basis for the work to proceed, the composing and arranging referred to above make up one integrated process. An exclusive feature of this category is that the time concept, manifested in periodicity, a clear metre, and other ‘musical logic’ as well as form based roles, is dissolved. All the other categories have a kind of linearity, that is, the song is written from A to B, or from A to A via B, none of which are features of VERTICAL 1B, in which the sounds and sections can be placed and mixed in any way.

**Vertical 2: Vertical composition, starting by defining the orchestra.** In this category, the composing originates from a tonally and stylistically comprehensive picture of what the finished composition is going to sound like. Consequently, the work is initiated by defining the orchestra. Expressed in terms of traditional composing, the work is begun by setting up the score: defining which instruments are going to be included in the ensemble for which the composition is intended. When the tonal picture has been set, the musical contents of the various parts emerge during the course of work, as described for VERTICAL 1A.

Even here the equipment provides the basis for the method of working by making possible the type of musical ‘thinking’ possessed by very practised composers, enabling them to write for an entire orchestra directly into the score, and simultaneously to ‘hear’ the sounding result while writing.

**Discussion**

The description given by Sloboda (1985) of the differences between the ways in which Beethoven and Mozart created their music have similarities with what have been described here as features of the horizontal and vertical ways of creating music. Whereas Beethoven worked out his themes and melodies over a long period of time, with ‘much compositional effort before anything was committed to paper’ (p. 112), and completing the melody line before adding the other instruments (horizontal composition), ‘Mozart scores show no evidence of this type of struggle.... Maybe [he has] control over the speed of “playback“ so that he is able to stem the flow of inspiration for long enough to adequately record all the details. In this case we would
expect to see a score written out "vertically" with each section completely transcribed before going to the next section' (p. 113).

Sloboda (1985) describes Stravinsky's way of composing as moving from initial conception to final theme over a short time span, and 'the time scale was short enough for the whole process to be kept in mind until the final version. The structure of the work – short, contrasting episodes – would help such a strategy. At one level, each section of 20 or 30 bars could be composed as a self-contained entity' (p. 112). This might indicate a way of working similar to the process described as the vertical categories.

The qualities and features of the different compositional categories can be compared and discussed in terms of the perspective of the creator in the process: in what position and role is the creator in relation to the music and the imagined others (i.e., co-creators, members of the band, audience) in the context of the composition?

In HORIZONTAL 1 and HORIZONTAL 2 the creator is both composer and musician, and as such remains at the centre of the process during composition, whereas the other musicians (represented by the computer) are accompanists. The various steps in the creation of a piece follow the same pattern: composing-rehearsing-performing (recording).

In VERTICAL 1 the creator has the role of a composer creating music by means of and in interaction with the computer. The function of being a musician is reduced to a minimum in that the technology is subservient to the mediating role of a musician. Instead, as a composer, the creator interacts with the sounds of the computer.

The way of working in VERTICAL 1A could be described as attributable to the equipment, whereas VERTICAL 1B has much in common with the ideas and ways of working at electro-acoustic music and sound compositions (e.g., Paynter & Aston, 1970).

In VERTICAL 2 the composer has the position of a conductor creating music in front of an orchestra, whose members execute the musical ideas the way they are instructed to. Although the equipment was a necessity for this way of working for the participants in this study, the method cannot be described as attributable to the fact that a computer was used, but rather the computer makes it possible for the 'untrained' composer to create music in a similar way as the most skilful composers have done through out history, writing the score for the whole orchestra directly.

A sequencer program involves a transformation of knowledge and skills used in traditional ways of creating music. Thus, the question may be asked as to how different software, not based upon traditional transfer of data via a keyboard, might affect the outcome with respect to ways of creating music, and how the computer is used. Whatever the prerequisites are, it may be assumed that there is a good deal of variation in the different ways of executing the tasks. It might be, though, that other types of software in which the creation of music is done in an unconventional way, for example by painting music with a midi-brush, would further diminish the effects of instrumental skill, and thus separate performance and composition to an even greater extent. Hence, such a program would offer a more equal point of departure, as it possibly would force all the participants to use the computer in a more exploratory way.

Although the technology to some extent mediates ideas of how to create music, the computer seems to have had the function of a tool for realising musical ideas, and thus being more or less transparent in the creative process. The 'transparency' of the computer might explain why reflections on the computer itself and its function, such as those discussed by Turkle (1984), were not observed in this study. According to Turkle, computers might also lead to an escape from the real objects to their representations. In music, though, it seems to be the reverse: computerised tools involve a shift from representations of music to the music itself.

Different instruments yield different results depending on their sound, construction, playing technique, and the stylistic associations they give rise to. The computer, though, seems to
operate on what might be described as a higher level in a ‘media-hierarchy’, that is, the different instruments and ways of expressing musical ideas are restricted to the potential of the computer. This is shown in that some of the participants used the computer equipment to do ‘drawing-sketches’, while others have chosen to ‘paint’ with sounds. This is not to say that the computer has no limits as a medium for composition, nor that it does not steer the way in which the music is created at all. However, the stylistic variety of the music in the present study, might indicate that the computer is not as controlling either for the ways of creating music, or stylistically, as could be expected according to the general view of the connection between computers and certain styles and genres of music. On the contrary, it may offer the freedom intended by the choice of the computer as medium.

In conclusion, all the participants succeeded in creating music. This is in no way a trivial finding, as it contradicts a well established conception of composition as being something that can only be done by a few, specially gifted individuals. The result might thus indicate that with computers bridging the gap between musical thinking and performance, every person should be able to create music (Sloboda, 1988, 1992), an idea confirmed in earlier studies using music technology for composition (e.g., Kratus, 1985, 1989; Folkestad, 1991).

However, our results show that the ways in which music is created varies between individuals, and between different kinds of music. An important implication of this is that school should not teach the method of composition, but rather create a context in which the pupils can explore their own ways into music composition. The knowledge of different strategies in composition which pupils spontaneously develop on their own, some of which has been presented above, should be of great value to teachers guiding their pupils into the adventures of musical creation.

Computers are likely to be increasingly used in music education at all ages, especially for creative activities like composition. Thus, further research in this field will become increasingly important. It has the potential to contribute valuable knowledge for teachers who implement and develop computer based musical activities in school.

References
Improvisation and Variation between Performance and Notation
by Heike Blumenberg, Berlin

I

Improvisation is certainly the most original way of musical practice and serves different purposes. Some works help the player to present his wealth of ideas and technical skills, others are elaborated compositions based upon a well planned concept even if they outwardly look like improvisations. Nowadays primarily popular among jazz musicians the art of improvisation can possibly be traced back to the medieval liturgical chant (D. Hiley, 1996, col. 541-553). In the course of the following centuries in written and non-written music developed a lot of improvisation techniques such as counterpoint ("contrapunto alla mente"), diminution, ornamentation, and variation. Since the sixteenth century the art of improvisation has not only been applied to vocal but also to instrumental music, so that three main types of improvising instrumental music can be distinguished:

- The first type comprises all kinds of free improvisation like fancies/fantasias, preludes, impromptus, musical meditations on pictures or texts/poems, cadenzas of solo concertos, the works of the "musique concrète" and aleatoric music.
- The second group comprises all kinds of variation. One kind is based upon small harmonic or melodic formulas, e.g. dances (Passacaglio, Chacona, Folia etc.) or grounds, the other one is based upon whole melodies, e.g. variation cycles or choral arrangements.
- The third group, the so called "guided improvisations", comprises all works where a player's ornamentation during performance is necessary (thorough bass compositions). For this study these works have been omitted because they mainly focus on only one improvising technique, the ornamentation (H. Meister, 1980, p. 355-368).

II

Terms like creativity, idea, invention, originality, and surprise are usually associated with the word "fantasia". The appearance of "fantasia" as a musical form changed over the years. Those fantasias that look like "frozen performances or frozen improvisations" developed in the middle of the eighteenth century (P. Schleuning, 1973, p. 86) when Carl Philipp Emanuel Bach published the chapter "Von der freyen Fantasie" in his "Versuch über die wahre Art das Clavier zu spielen" (1753/1762). While eighteenth century fantasias were characterized by the loose joining of different musical motifs, high virtuosity dominated the nineteenth century works of e.g. Frédéric Chopin or Franz Liszt. But both types of fantasia were connected
through the idea of uniqueness, originality and genius referring to the inspired composer having an unlimited reservoir of ideas as well as to the brilliant player having an unlimited reservoir of technical skills.

Originality and uniqueness linked with breaking rules, however, had not always been seen as a proof of genius. In earlier centuries disregarding the composition rules proved a composer's ignorance and incompetence. Thus elaborated polyphonic structures - taken from contemporary vocal music - characterized many early Spanish vihuela fantasias (e.g. Enríques de Valderrábano; J. A. Griffiths, 1984, p. 254). Other works, especially those ones of Luis de Milán, are more improvising compositions by dint of using only one idea or motif woven into different musical contexts. Nevertheless, the early vihuela fantasias had intrinsic musical structures that required the player's and listener's concentration and caused a quiet and introvert atmosphere during performance (J. A. Griffiths, 1984, p. 538). This was the very opposite of what later fantasias did, which was to arouse the listener's admiration and impression of the player's extrovert, or sometimes excessive, performance.

A more reflective improvisation which, however, uses the means of virtuosity and symphonic volume of sound is Marcel Dupré's organ work "Le chemin de la croix". The original version of this composition was an improvisation on Paul Claudel's fourteen contemplative lyrics of the crucifixion given by Dupré in 1931 at the Conservatory of Brussels. For this composition Dupré chose a fund of characteristic musical motifs (which can be compared to Wagner's "Leitmotive") to illustrate each station. The music of these fourteen short pieces commutes between reflexion and drama depending on a station's particular subject. It is improvisation linked with high demands regarding contents and technical skills that dominates these works, not in an extrovert and virtuoso but in an introvert and reflective manner.

III

Improvisation on a musical theme or a harmonic or melodic formula became very popular through the Chacona and Passacaglio variations of sixteenth century Spanish lute music and later through Italian and English keyboard music (R. Hudson, 1981, p. 169-176). Many of these compositions like e.g. "Divisions upon a ground" have something of extempore performance about them. They all used diminution as the main improvisation technique. The great number of contemporary printed methods of diminution (e.g. Ganassi, della Casa, Ortiz etc.) proves the high rank of diminution among sixteenth century instrumental improvisation techniques.
The technique of varying a melody by diminishing its rhythmic values from variation to variation can be followed through the following centuries. Johann Sebastian Bach also used diminution as a variation technique in his famous organ Passacaglia in C minor, but he combined it with a special harmonic development: at the end of each part of this work, Passacaglia and Thema fugatum, Bach reduced the harmonic formula to two chords: C minor and G major. A comparable work to this great Passacaglia was written about two hundred years later by Johann Gabriel Rheinberger. His impressive Passacaglia in E minor - the last movement of his organ sonata No. 8, op. 132 - shows the methods of diminution and variation very clearly by its exhausting all possibilities to diminish the rhythmic values and to harmonize the Passacaglia theme. Compared to each other both works have different concepts: Rheinberger elaborated brilliantly the technique of diminution and harmonic variation, while Bach added a second idea to the method of diminution by reducing the harmonic formula to two chords so that the work ends with a composed harmonic standstill.

A further concept of variation can be noticed in Franz Schubert's variations on the lied "Trockene Blumen" for flute and piano. The mere thought that the theme is taken from the lied-cycle "Die schöne Müllerin" gives rise to the idea that Schubert composed a dramatic scene by summarizing the story told in the lyrics of "Die schöne Müllerin". Besides the non-musical concept keeping this work together Schubert used all variation techniques customary at the time: diminution, imitation, ornamentation and the character variation, where all musical parameters change simultaneously.

IV

Improvising music which has been written down can serve different purposes. Some works function as memory aids or "transcripts" of once played performances (fantasias) or as examples to stimulate the player to invent further diminutions and variations (grounds). Invention combined with technical brilliance is the predominant idea of these works.

Other works like the early Spanish vihuela fantasias are well elaborated imitations or counterpoints, where improvisation only refers to a work's musical form and to the manner in which its theme or main motif has been chosen.

Moreover there are works which are elaborated compositions, where diminution and variation are only used as a composition's framework. In these cases the improvising character has been replaced by a superordinate idea that keeps all parts of the composition together. This idea can refer to music (Bach, Passacaglia) or to something outside music (Schubert, variations).
These works are better planned and more structured compositions than extempore improvisations, because the composer uses improvisation or variation only as a means of expressing an idea beyond performance. And finally, the listener's perception and first impression of these compositions can differ entirely from the reader's observations.

References

Tonal Dynamics in Jazz Improvisation

Topi Järvinen
University of Jyväskylä
Department of Musicology

Improvisation is an interesting musical activity, because it imposes strict temporal constrains on the creation of music. For while improvising one cannot go back and revise some past event that in retrospect appears unsuitable in conjunction with the other events in the improvisation. In two recent studies (Järvinen 1995, 1996) the statistical distribution of tones was analyzed in numerous bebop-styled jazz improvisations based on the Rhythm Changes chord progression. Specifically, it was examined how the improvisers used tonality and meter in order to cope with the cognitive strains present in real time music making process. In the previous studies the treatment of the data was for the most part atemporal. In this study by using partly different body of materials I have investigated how the tonality changes in the course of the Rhythm Changes chord progression. The main focus of interest was on the interplay of local and global tonality.

The forty two chorus length improvisations were analyzed by investigating the statistical distribution of the tones. In addition to the total frequency, also the frequency of tones in each measure and on each metrical position was examined. The obtained local tone-frequency profiles were compared with the acquired global profile and a set of 44 calculated chord profiles shown in Table 1 (pitch-class saliences according to Parncutt, 1988). The aim was to choose the ones that jazz theory usually predicts as the most probable substitutions (e.g., Levine 1995).

TABLE 1. The chords that were used as comparisons in the chord progression analysis.

<table>
<thead>
<tr>
<th>chord type</th>
<th>chords</th>
</tr>
</thead>
<tbody>
<tr>
<td>M7</td>
<td>CM7, FM7, DM7, GM7, CM9, FM9</td>
</tr>
<tr>
<td>dom7</td>
<td>G7, A7, C7, D7, E7, Bb7, Bb7b5, Eb7b5, Ab7b5, Db7b5, B7, E9, A9, D9, G9, C9</td>
</tr>
<tr>
<td>m7</td>
<td>Am7, Dm7, Gm7, Fm7, Bm7, Em7, Dm9, Am9, Em9, Bm9</td>
</tr>
<tr>
<td>diminished</td>
<td>F#°7, G#°7, C#°7</td>
</tr>
<tr>
<td>half-</td>
<td>Bm7b5, Em7b5, G#m7b5, C#m7b5, F#m7b5, D#m7b5,</td>
</tr>
<tr>
<td>diminished</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1 shows the correlations between the local tone-frequency profiles (loc.chord) and both the metrically weighted global profile (wt.glob.prof) and the calculated chord profiles (calc.chord) based on the original progression. The graph indicates that the correlation between the local profile and the global tonality is higher than the correlation between the local profile...
and the calculated profile for almost every chord. There are, however, two notable exceptions. In the latter part of the fifth measure the importance of the global tonality drops and at the same time there seems to be more emphasis on the local tonality of the chord.

FIGURE 1. The correlations between the local tone-frequency profile (loc.chord) and both the weighted tone-frequency profile (wt.glob.prof.) and the calculated chord profile (calc.chord) based on the original progression.

FIGURE 2. The correlations between the local tone-frequency profile and both the weighted tone-frequency profile and the calculated chord profile.
These results suggest that most of the time the global tonality is the most important tonal reference point. The problem is, however, that the musicians may have used chord substitutions, and thus the results give distorted view of the improvisers preferences. For that reason, it was also examined if the players had used some substitutions in place of the basic chords used in Figure 1. Each of the chords in listed in Table 1 were correlated with each local tone-frequency profile, and the chord that yielded the highest correlation was chosen in each case. Based on the results a new chord progression was constructed (see Fig. 2). The correlations for the chords are higher than in Figure 1. According to these results the CM9 chord is used extensively: ten out of the sixteen chords is a C major chord. In this graph the importance of the F major chord and the preceding C dominant chord is even more unambiguous than in Figure 2. It should be noted that the data that was used in the above analysis was constructed by averaging all the A-sections. However, the same general principles can be found in the individual A-sections.

In order to investigate how diffused the local profiles are, standard deviation values were calculated for each profile (see Fig. 3). The result indicate that the profiles for the C major (first and seventh measure), F major chord (sixth measure), and C dominant chord have relatively sharp focus compared to the other local profiles. In other words, the musicians have clearly favored some tones over the others. Moreover, it is interesting that the standard deviation is higher for the profiles in the first half of the measure and lower in the second half of each measure. Thus, a pattern emerges which is similar to the one that can be found in the metrical structure, namely the alternation of strong (even) and weak (odd) parts of the measure/beat. The most prominent exception of this pattern is in the fifth measure where the C9 chord gets a relatively high standard deviation. This result is, however, well in concordance with the earlier findings. For the C9 chord has an important function in the chord progression as the secondary dominant for the F major chord.

![Graph showing standard deviations for weighted average A-section tone-frequency profile.](image)

To investigate the development of this progression further, tone-frequency profiles were obtained also for the B-section of the Rhythm Changes chord progression. Figure 4 shows the correlation of the local profile with the global A-section tone-frequency profile (filled ball), with the most similar chord the in the set of calculated chord profiles (diamond), and with the diminished chord one half-step above the original chords root (empty ball). The chords on the x-axis display the calculated chord profile which has the highest correlation with the tone-frequency profile in that measure. Figure 4 shows that musicians seem to favor chords that extend beyond the basic triad and seventh chord. Moreover, although the correlation for the diminished seventh chords is never the highest in the set, the graph indicates that it is an important element in the progression. The results show that the correlations with the appropriate diminished chord are always the highest right before the chord changes. In other word, in the even measures there seems to be tendency for the musicians to prepare for the next chord by using tension (cf., Järvinen 1996).

![Figure 4](image)

FIGURE 4. The development of tonality in the B-section of the Rhythm Changes chord progression. See text for further explanations.

The relationship of the global A-section profile and the local profiles of the B-section is interesting. The negative correlations in the beginning of the section are surprising: it may imply that the musicians want to emphasize the beginning of this new section by avoiding the tonality of the previous section. The relatively high correlation between the global A-section profile and the local tone-frequency profile in the sixth measure seems to suggest that the musicians may start to anticipate the tonality of the following A-section. This seems to be a reasonable assumption, for the original underlying chord D dominant is the secondary dominant for the dominant chord of C major, and therefore it is fairly closely related to key of the A-section.
FIGURE 5. Standard deviations for the tone-frequency profiles the B-section.

The standard deviations were also obtained from the B-section (see Fig. 5). They indicate that in most measures the tone-frequency profiles are more diffused in the second half than in the first half. In other words, musicians place different amount of emphasis on the chords depending on the position within the measure.

This study replicated the earlier results (Järvinen 1995, 1996), but it offered also some important new additions and clarifications. These results seem to suggest that at least in this type of music both global and local tonal orientation affect the process of improvisation. These present results seem to indicate that both the process of improvisation and the notion of tonality is very dynamic. It was concluded previously that, for example, in the A-section there are reference points or chords which are outlined more carefully than the rest of the chords. While this is essentially correct, it implies too static a view of jazz improvisation, because it seems that the musicians are constantly thinking forwards. In other words, the structurally important chords are prepared in advance which further underlines their significance. It is important to note, however, that this does not mean that there is no fairly static global tonality. On the contrary, the results also show that the musicians seem to use also some type of tonal hierarchy as a cognitive reference point in their improvisations.

References


Music Performance: Representation, Practice, and Reading
Expert Pianists' Mental Representations: Evidence from Successful Adaptation to Unexpected Performance Demands

Andreas C. Lehmann & K. Anders Ericsson, Florida State University

The production of music by solo performers often appears so effortless that the audience is led to believe that its generation must be automatic. Contrary to this impression, we know that the musicians' performance is mediated by complex cognitive processes (see Palmer, 1997, for a review). For example, musicians need to monitor their own performance, anticipate and compensate for unexpected memory problems, make slight expressive adjustments in response to the instrument or the acoustics, and interact and synchronize with other musicians. Thus, musicians have to acquire mechanisms that will allow them to perform reliably while still being able to make those adaptive changes under time-critical conditions.

Experts in various domains are found to adapt maximally to the demands of representative tasks in their respective domains of expertise, and it is commonly believed that this domain-specific superiority of expert performance is due to increased automaticity and reduced cognitive control. However, a recent review (Ericsson & Lehmann, 1996) found that expert performers in many domains of expertise had acquired complex mental representations that resulted in higher accuracy, more control, and greater flexibility of performance compared to lesser experts. In the present study we apply these findings to music performance and investigate individual differences in expert musicians' ability to form and effectively use mental representations (also Ericsson, in press). In particular, we are interested in the speed at which performers can memorize short pieces of music to fixed criteria of tempo and accuracy, and how accurate they are in performing those pieces of music under different conditions that require varying degrees of flexibility. For example, performing a piece at a different tempo should be easier than playing only selected measures or transposing it to a different key. The discussion will place our findings regarding the role of mental representations of memorized music into the larger context of research on expert performance and deliberate practice.

Previous studies on memorization have not always controlled for exposure to the material, practice time, and strategy of memorization (e.g., Rubin-Rabson, 1941; Nuki, 1984), thus making it difficult to study individual differences in subjects' ability to memorize. In the present study we control for speed of performance and amount of exposure to the stimulus material during memorization by using the "accompanying paradigm". This method, where expert pianists play along with a pre-recorded solo part was previous employed in a study of piano sight-reading to simulate the representative situation that is most commonly associated with sight-reading, namely the accompaniment of a soloist by a pianist (Lehmann & Ericsson, 1993).

To capture the structure of expert performance, Ericsson and Smith (1991) advocate the use of tasks that are representative for the domain of expertise. Therefore, to assess the pianists' mental representations once the music was memorized, experimental tasks were designed that required the pianist to perform the music under very different conditions--many of which are mentioned as practice strategies in the piano teaching literature (e.g., Uszler, Gordon, & Mach, 1995). For example, in one condition the pianists played with only one of their hands, in another condition they played according to structural boundaries, and in yet another condition the pianists transposed the memorized music to a different key. At a more general level, the tasks challenged the pianists' long-term working memory skills by forcing them to mentally access and manipulate encoded bimanual information, structural information about the piece, motor programs, and detailed pitch information.

The purpose of this study was to measure individual differences among expert pianists with regard to their ability to memorize unfamiliar music, and to assess their flexibility when forced to perform the memorized music under varying task demands. Finally, the ability to memorize was related to
other indicators of music performance skill to gain information about the acquisition of this specialized skill.

**Methods**

**Subjects.** Subjects in this experiment were 16 expert pianists from a school of music, of which all had participated in a previous study on sight-reading (Lehmann & Ericsson, 1993).

**Stimuli.** The stimulus material consisted of two 8-measure segments from Franz Schubert's sonata for violin and piano (op. 137, D 384). The order of presentation for the segments was counterbalanced among subjects.

**Procedure.** In each of two consecutive sessions pianists memorized one of the two segments. Two paced, note-perfect renditions from memory were the required criterion for mastery. Each learning trial in the memorization phase consisted of two parts (see Fig. 1, left panel). During the first part, the pianists played the music from a computer display, paced by a prerecorded solo part heard over loudspeakers. During the second part of the trial, the pianists heard the solo part again but had to recall the accompaniment from memory. A metronome track with clicks preceded the solo part in order to facilitate the pianists' entrance. After successful completion of the learning phase, subjects were asked to perform several types of reproduction tasks which were also paced (see Fig. 1, right panel).

**Figure 1.** Design of learning trials (left) and experimental performance tasks (right).

**Learning Trial**

- **Pacing solo voice**
  - Metronome clicks
  - Subject plays from music notation on computer screen

**Experimental Task**

- **Pacing solo voice**
  - Metronome clicks
  - Subject plays from memory

**Experimental tasks for performance under varying conditions.** The expert pianists had to play the memorized segment at both 166% (fast) and 50% (slow) of the original tempo encountered during memorization. Next, they reproduced selected parts of the memorized segment at the original tempo set by the pacing voice (i.e., right hand only, left hand only, odd-numbered measures only, even-numbered measures only). Also, the segment had to be transposed from its original key of F to F#. Because the music had to be performed from memory for all tasks, subjects intermittently performed the original version to ensure that they still had an intact representation of the original piece. Failure to perform the piece to criterion resulted in additional learning trials.

**Collection and analysis of the data.** Performance data (MIDI) from the recall phase of the learning trials and the reproduction tasks was recorded with a computer connected to a Yamaha KX88 keyboard, on which the subjects performed. In addition, a video recording was made of the pianists' hands. Performance data from the experimental tasks was converted to ASCII text and scored for accuracy of pitch and timing. The accuracy of the two temporally varied reproductions were aggregated into a single score, as were the accuracy scores of the different selective reproductions. However, the accuracy for the transposition task was analyzed separately. Also, the fingerings that subjects had chosen during performance were encoded from the video (for the
Individual differences in subjects' ability to memorize the short music excerpts will be examined first. Next, performance under changing task conditions will be analyzed, first the tempo changes, second the selective reproduction, and finally the transposition. After that, the ability to memorize and perform flexibly under varying task conditions will be related to various indicators of skill acquisition and performance.

The average number of trials (mₚ) required to play the piece perfectly from memory on two consecutive trials was 9.5 (SD = 4.63) for Segment 1 and 11.56 (SD = 5.15) for Segment 2. Our measure of memory ability will be mₚ, because the best memory performance is associated with the smallest number of trials. Speed of memorization for the two segments was reliably correlated (r = .76, p ≤ .001), suggesting a general ability to memorize music (see Fig. 2). Further evidence for a general memorization ability was given by a reliable correlation (r = .53, p ≤ .05) between memorization ability in this study and a test of memory which had been administered as part of an earlier study of sight-reading (data from Lehmann & Ericsson, submitted).

As expected, subjects were highly accurate in reproducing the pitches of the memorized piece at different tempi and while playing selected parts (see Fig. 3)—even when each keystroke had to fall within +/- 100 ms of its specified time to be scored as accurate. Correlations between performance for the two pieces did not reach significance, possibly due to a restriction of range. In the following, the score on the temporal variation and selective reproduction test will be combined to form a single score for adaptive reproduction.

Unlike the other recall tasks, which were all performed with a high overall degree of accuracy, transposition seemed to be more complex and resulted in reliably lower performance accuracy (see Fig. 3). The correlation between error and time corrected accuracy of transposition among the two pieces was significant, r(13) = .75, p ≤ .01, suggesting that our measure of generative reproduction was reliable. To generate an optimal mapping of fingers to keys (fingering), subjects had to respond to the spatial change and anticipate a following chord which required a certain finger
combination. We specifically investigated the first two measures of the left hand for Segment 1, which consists of a complete reversal of black and white keys in the current transposition. Choosing a fingering that reflected those constraints was reliably associated with higher performance accuracy (corrected for error and timing), $r = .60$, $p \leq .05$.

Are speed of memorization, adaptive reproduction and generative reproduction mediated by the same underlying representations? The correlations between these abilities are uniformly high, suggesting a shared representation that possibly mediates all these performances (see Tab. 1). How do these abilities and their mediating representations relate to other indicators of musical skill? Neither a direct measure of sight-reading ability nor accumulated piano practice were reliably related to memory ability (see Tab. 2). The only reliable relation was found between adaptive reproduction and amount of accompanying experience.

**Table 1. Correlations between ability to memorize and memorized performance under varied conditions.**

<table>
<thead>
<tr>
<th>Ability to Memorize</th>
<th>Adaptive Reproduction</th>
<th>Generative Reproduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptive Reproduction</td>
<td>.70 **</td>
<td></td>
</tr>
<tr>
<td>Generative Reproduction</td>
<td>.75 **</td>
<td>.72 **</td>
</tr>
</tbody>
</table>

* p < .05; ** p < .01

**Table 2. Correlations between memorized performance and indicators of skill acquisition.**

<table>
<thead>
<tr>
<th>Ability to Memorize</th>
<th>Adaptive Reproduction</th>
<th>Generative Reproduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accumulated Deliberate Practice</td>
<td>.37</td>
<td>.19</td>
</tr>
<tr>
<td>Accompanying Ability</td>
<td>.25</td>
<td>.40</td>
</tr>
<tr>
<td>Accompanying Experience</td>
<td>.28</td>
<td>.54 *</td>
</tr>
</tbody>
</table>

* p < .05; ** p < .01

**Discussion**

When expert pianists acquire a musical performance at a specific tempo, their performance is mediated by mental representations that allow the musician to adapt and reproduce the music accurately (notes and timing) under a range of different performance conditions. More importantly, the data reveal evidence for individual differences even among expert pianists in their ability to construct and use these representations. Correct performance at different tempi and selective reproductions of parts of the music were closely linked to the ability to generate new motor programs during transposition. The two measures of adaptive reproduction were closely related to the speed of original memorization, which in turn was related to memory performance in a sight-reading context (data from Lebmann & Ericsson, submitted). This relation between domain-specific memory ability and the quality of representation which mediates flexibility and control is observed in many types of expert performance.

The actual structure of these representations and how they mediate adaptive and generative reproduction is currently under investigation. We are presently analyzing process traces consisting of trial-by-trial retrospective verbal reports provided by the pianists in this study during the memorization phase as well as retrospective reports from the reproductive tasks. Also, an ongoing analysis of the errors committed during transposition investigates their relation to mental representations. How is flexibility in performance acquired? Studies of expert musicians have
shown that expert performance in a given domain is not one single general skill, but rather a combination of specific subskills acquired to meet the demands of particular performance situations, such as playing unfamiliar music at first sight or performing well-rehearsed repertoire (also Lehmann, in press). Earlier research has shown that sight-reading ability is related to the accumulated amount of accompanying (Lehmann & Ericsson, submitted) and that level of solo performance of repertoire is related to accumulated amount of deliberate practice (Ericsson, Krampe, & Tesch-Römer, 1993). The moderate association between adaptive reproduction and accompanying experience is not entirely surprising, given that accompanying involves performance in situations where adjustments of tempo and rapid responses to unexpected problems frequently occur. On the whole, the ability to rapidly memorize new music and the associated ability of its generative reproduction among expert musicians were not found to be related to sight-reading or well-rehearsed performance. Thus, it must likely reflect yet another acquired subskill and its associated mental representations.

Recent research on expert performance (see Ericsson & Lehmann, 1996, for a review) reveals the central role of integrated mental representations that allow experts to acquire new skills, knowledge, and repertoire. Those representations also enable the experts to monitor their performance and respond flexibly to problems and changes in external conditions. Studying the structure and acquisition of mental representations associated music is important for understanding mediating mechanisms of expert performance and for improving the training of future experts.

References


Demonstrating Inner Hearing Among Musicians

WARREN BRODSKY, BEHAVIORAL SCIENCES, BEN-GURION UNIVERSITY OF THE NEGEV, ISRAEL
AVISHAI HENIK, BEHAVIORAL SCIENCES, BEN-GURION UNIVERSITY OF THE NEGEV, ISRAEL

Imagery has long been an important aspect of psychological inquiry. Empirical investigations have mostly involved the visual modality. However, the auditory modality has currently become the focus of experimental investigations. Researchers have assumed that auditory imagery is related to audition in the same way that visual imagery is related to vision. Anecdotal evidence has suggested that auditory images or "sounds in the head" are just as vivid as visual images or "pictures in the head." Recently, studies have not only demonstrated the quantification of auditory images, but pointed to clear evidence that this representation codes the extension temporally, unfolds in real time, and has strong links to the melodic and dynamic components of sound. Many researchers have identified musical imagery as a special case of auditory imagery as this particular imagery involves explicitly musical stimuli. Moreover, that a unique subtype of musical imagery is that which is exclusively cued from musical notation; a phenomenon known as inner hearing (also referred to as 'notational audiation' or 'mental score').

For the majority of ordinary people, music may be primarily an experience of the auditory mode. That is, music is something that is heard. Yet, musicians rely on musical imagery and richness of qualia to inspire and guide performances, just as much (if not more) than its actual external sounds. It is believed that musicians create music by hearing-it-out as opposed to picking-it-out on an instrument, and that musicians do not merely see the score - they hear it in their 'mind's ear.' However, these premises continue to be speculative. Empirical cognitive investigations have not as yet demonstrated inner hearing of a score independently to the visual memory processes involved. This paper will present the early findings of a research study currently in progress on the inner hearing of musicians. The study highlights a cognitive paradigm by which abilities of inner hearing can be demonstrated or refuted. As a final goal, the study anticipates investigating some questions on the nature of inner hearing, such as: (1) the development of inner hearing as pertains to age, experience, exposure, education, and training; (2) the inter-relationships between inner hearing, sight reading, and musical literacy; and (3) differences in ability to generate musical images from the printed page among musicians subgrouped on the basis of their chosen instrument (strings, wind, piano, etc.), as well as according to their mode of musical activity (conductors, composers, performers, teachers).
Introduction

Is instrumental practice a planned activity? Do all instrumental students plan their practice? If they plan, how do they conceive planning? These are some of the questions that set off the research reported in this paper, where I focus on higher instrumental students’ planning of instrumental practice. I am not concerned with performance plans and the development of mental representations of performances (see Sloboda, 1982, Gabrielsson (in press), Lehmann, 1997), but with the planning of practice activities.

My assumption is that all students do some planning in connection with practising, because they believe that planning of some sort is important for the efficiency of learning. They may also use planning to meet immediate personal needs, and for other reasons. The educational implication of this is that if we have a better understanding of students’ practice planning, we may be in a position to help them utilise planning for more efficient practising.

Since the publication in 1960 by Miller, Galanter and Pribram of their “Plans and the structure of behavior”, there has been a resurgence of interest in research on planning. Today, there are several possible theoretical frames of reference for the study of students’ practice planning. My frame is theories of teaching, or didactic theories. I regard practice as a teaching activity, where the practising student has to behave like a teacher. Most of the research on teacher planning in the last 10-20 years has been carried out with cognitive theory as the frame of reference. “Teacher thinking” has emerged as the overarching concept, and the ultimate goal is “to construct a portrayal of the cognitive psychology of teaching...” (Clark and Peterson, 1986, p.255). This frame of reference has supplemented the process-product approach to the study of teaching effectiveness, where the relationship between teacher behavior and student achievement is studied. Believing that teachers’ actions to a great degree are caused by teachers’ thought processes, “to understand teacher planning is to understand how teachers transform and interpret knowledge, formulate intentions, and act from that knowledge and intentions” (Clark and Dunn, 1991, p.184). Now, for my purpose, the concept “teacher” and “teaching” can be replaced by “student” and “practice”, giving my research a theoretical foundation similar to the teacher thinking research.

Research questions and methods

My research context is an academy of music, with students in performance (instrumental and vocal), church music (with organ as major instrument) and music education departments, organised in a 4-year undergraduate program.
There is no previous research on instrumental practice planning, and I regard my research as explorative, trying to give some first impressions from a complicated field of enquiry. For this purpose, my broad research theme is:

*Is instrumental practice a planned activity?*

This question will be explored with both a quantitative and a qualitative approach. My paper will report preliminary results from a part of a larger project, involving questionnaires to the student population in spring 1991, spring 1995 and spring 1996, as well as interviews with students in the spring of 1995. I will concentrate this paper on empirical information from the 1995 part of the project. Here, 145 of 167 students (87%) answered a questionnaire about several aspects of practising, among them were also two questions on planning. Furthermore, I interviewed 10 students about their practice, also about practice planning. I will combine information from these two sources in this paper.

In the questionnaire, I presented students with two fixed-alternative questions on planning, and asked them to comment on their answers in corresponding free-response questions. The two fixed-alternative questions were concerned with *when* and *how often* they plan, and if they regard themselves as *systematic users of planning* or not. 75 of the students gave a total of 107 free-response comments to the two questions. These comments, and the 10 interviews, gave information on a broad area of aspects of planning. In my analysis, I am influenced by grounded theory (Glaser, 1992) in creating categories from the written and spoken information supplied by the informants. The presentation of results will, accordingly, be centred around emerging themes and categories. The quantitative information from the two fixed-alternative questions will be placed under the appropriate theme.

### Results

*Do students plan?* The *occurrence of planning* is exemplified by comments ranging from “Planning is Alpha and Omega” to “I practise totally without a plan”. It seems that most, may be all students, are “planners” of some sort. Even the student saying that he never plans, did not answer with the “never” categories in the question on when they plan (see below).

The *time period for planning* was asked for in the first fixed-alternative question, and this theme also got several free comments. The questionnaire gave the following instruction: “When you are practising, you will have to decide on what you will practise, how you are going to practise, and what you want to accomplish. All of this has do with *planning* of practice. This planning may be carried out both before, during and after practising, and it may be something you spend a lot of time on, or only decide upon at the moment. I have singled out five time periods and ask: How often do you plan and make decisions on these points of time?” The time periods were: “Before a practice day”; “In the beginning of a practice session”; “During practising”; “Shortly after practising, to prepare for later practising”; and “Between practice days and practice sessions, in your “Spare time””. Table 1 shows that all time periods were represented with planning activity. Planning *before a practice session* is the type of planning most students are engaged in, here 76% report that they “always” or “often” do some planning. 49% of the students report that they “always” or “often” plan *before a practice day*. Planning shortly after practice sessions, or in their “spare time” between practice days and practice sessions is much less used by the students.
Table 1
Time periods for and occurrence of planning

<table>
<thead>
<tr>
<th>Time period</th>
<th>Always</th>
<th>Often</th>
<th>Sometimes</th>
<th>Seldom</th>
<th>Never</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Before a practice day</td>
<td>19</td>
<td>14</td>
<td>49</td>
<td>35</td>
<td>22</td>
<td>16</td>
</tr>
<tr>
<td>In the beginning of a practice session</td>
<td>29</td>
<td>21</td>
<td>76</td>
<td>55</td>
<td>21</td>
<td>15</td>
</tr>
<tr>
<td>During practising</td>
<td>17</td>
<td>12</td>
<td>51</td>
<td>37</td>
<td>43</td>
<td>31</td>
</tr>
<tr>
<td>Shortly after practising</td>
<td>5</td>
<td>4</td>
<td>23</td>
<td>17</td>
<td>43</td>
<td>32</td>
</tr>
<tr>
<td>Between practice days and sessions.</td>
<td>3</td>
<td>2</td>
<td>36</td>
<td>26</td>
<td>44</td>
<td>32</td>
</tr>
</tbody>
</table>

There were marked individual differences in planning profiles between the students. A (statistical) analysis may reveal that the students can be distributed in major planning types regarding time period and occurrence of planning.

The time perspective of plans differ between students. Some make long-term plans, like this student: “I make practice programs with themes for every practice session for may be a month in advance. Then I use the first part of practice sessions to decide on exercises that fit my theme for this session”. Others are what might be called short-term or even on-the-spot-planners. This last category apparently has many supporters, judging from the number of comments like "During practising I feel my mood, physics etc., and then it is easier to decide what to do”.

The focus of planning tells us what the students are concentrating their planning effort on. Here, I discern the three following categories: Type of activity; the order of appearance of activities; and type of content in activities in a practice session. In “type of activity”, what they are doing is related to a purpose (like warming up, developing technique, and developing a repertory). With “order of appearance of activities” I mean the order in which the activities are practised (for instance with warming-up exercises first, then technical exercises, after that repertory). And with “type of content” I mean the physical and musical content in these activities (breathing exercises, scales, etudes, solo repertory, orchestra studies etc.). Some students plan in one or all of these focus categories, making alterations from week to week or from day to day. For instance, altering their warming-up exercises as a consequence of evaluation and planning. On the other hand, many students report that they have routines that simplify planning. Many students remark that they have routines for all or some of the three focus categories, ranging from “My warming-up program is quite firmly established”; and “I have a system of exercises, scales etc. which are used every day”, to “I have a daily program of technical exercises I go through every day. The rest of the day is quite impulsive”, and “I have a relatively firm order of appearance for the activities I go through (warming up, scales, etudes, solo repertory, chamber music etc.) so my planning is mainly to decide on what needs most attention today.” These comments, and others, also show that routines are dominating the practice time for some students, whereas routines are used in diverse connection with planning or more spontaneous activities by other students.

A special category of routines is established by the homework the students are given for their next lesson with their teacher. They meet their instrumental teacher once a week, and one
student writes that “I have my homework from my instrumental teacher. Consequently I have control over what I have to practise this week” This is, of course, the situation for most of the students, but some of them seem to rely more on this assigned program than others, thereby giving less thought to planning. This category might be called the-homework-does-the-planning, the students simply “go on practising”.

“I have always been careful to set up a plan (in my head) before a practice session” is a comment from several students. This points to the differences between mental plans and written plans, categories in a planning medium theme. The prevailing attitude seems to be that written plans are a nuisance, at least something to avoid, or something which “is nothing for me”. Written plans are mostly regarded as obstacles, preventing students from reacting to the unforeseen. From the written comments and from the interview information, my impression is that very few students utilise written plans. What is more common, is writing small comments in the score. This might at least be regarded as the outcome of some sort of planning.

The differences between outline plans and detailed plans are also commented upon, illustrating the elaboration of plans. One student combines these: “My planning is on several levels. I may have an outline plan for what I want to practise on a day, but the detailed planning is mostly carried out immediately before a practice session.”

Many students commented on the functions of planning: “To have efficiency. If I had no plan, I would be sitting in the practice room not knowing what I was actually doing, and then, consequently, I would spend a lot of time with no results.” This planning for efficiency seems to be the prevailing motivation, even if there are also indications that planning sometimes serves to meet personal needs (for instance “I have to plan because I take pride in being well prepared for my next instrumental lesson”).

There are also many students who comment on planning in a self-reproachful manner: “My planning has become better, but it may still improve a lot...”, “I am not good enough at planning” and “I am sure I could have made better plans” are all examples of this attitude. These self-evaluations and comments on the necessity of planning might be related to my second quantitative question: “Do you regard yourself as one who systematically uses planning when practising?” 144 of the 167 students answered this question, and the answers were distributed as follows on the fixed alternatives: 7% regarded themselves as “very systematic users of planning”; 19% were between “very systematic” and “medium”; 49% were “medium”; 20% were between “medium” and “very unsystematic”; and 6% were “very unsystematic”. From this I conclude that the students are arranged in an approximate normal distribution when they are asked to describe themselves as systematic or unsystematic users of planning.

Summary and discussion

The presented themes and categories for practice planning correspond to similar categories developed in research on teacher planning (see reviews in Clark and Peterson, 1986; Borko and Niles, 1987; and Clark and Dunn, 1991), and is at least partly an affirmative answer to the question posed as my research theme. The planning themes and categories presented here must, however, not be regarded as a complete and exhaustive list, but as an identification of important themes regarding students’ practice planning. There certainly are nuances within the
categories, and additional categories. My hope is that the themes and categories presented here may serve as a starting point for further research on practice planning.

The planning themes and categories identified in the students’ written or spoken information are summarised in table 2.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occurrence of planning</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Time period for planning</td>
<td>Before a practice day</td>
</tr>
<tr>
<td></td>
<td>In the beginning of a practice session</td>
</tr>
<tr>
<td></td>
<td>During practising</td>
</tr>
<tr>
<td></td>
<td>Shortly after practising</td>
</tr>
<tr>
<td></td>
<td>Between practice days and sessions</td>
</tr>
<tr>
<td>Time perspective</td>
<td>Long-term</td>
</tr>
<tr>
<td></td>
<td>Short-term</td>
</tr>
<tr>
<td></td>
<td>On the spot</td>
</tr>
<tr>
<td>Focus of planning</td>
<td>Type of activity</td>
</tr>
<tr>
<td></td>
<td>Order of appearance of activities</td>
</tr>
<tr>
<td></td>
<td>Type of content in activities</td>
</tr>
<tr>
<td></td>
<td>Home-work as plan</td>
</tr>
<tr>
<td>Routines (as above)</td>
<td>Mental plans</td>
</tr>
<tr>
<td>Planning medium</td>
<td>Written plans</td>
</tr>
<tr>
<td>Elaboration of plan</td>
<td>Outline plan</td>
</tr>
<tr>
<td></td>
<td>Detailed plan</td>
</tr>
</tbody>
</table>

One of the most striking impressions I gained, analysing students’ comments, was the diversity and great range of individual differences in attitude toward planning. This must also be accounted for in future research.

The interactive nature of planning illustrated in teacher planning research (for instance the interaction between type of content and time perspective) is also illustrated among the practice students. “Planning is a continuous process where the different activities of playing and practising are planned at different time periods,” writes one student. Many of the attitudes to planning among teachers also correspond to attitudes shown by the practising students. For instance the scepticism toward long-term and detailed planning, where the more experienced teachers rely heavily on outline plans, while novice teachers utilise more detailed plans. This difference between novice and expert teachers is, of course, at least partly an effect of the expert’s greater experience and subject matter knowledge. Is this also the case among practising students? This is a very interesting question for further research.

A crucial point for our understanding of (and research on) practice planning is the role of routines. Are routines a type of planning, or the result of planning? Or, are routines objects of planning? In teacher research, opinions are divided in this matter. The students responding with written and spoken comments in my research, mostly seemed to include routines as a
type of planning. However, some of them quite explicitly stated that they had very little need for planning, since they had established routines for many aspects of practising. This may account for the reliance on on-the-spot-planning among some of the students.

The *unpredictable nature of practising* is commented on by many of the students. And some of them obviously support the more intuitive, on-the-spot type of decision making as a general approach to planning, sometimes combined with discouraging descriptions of fellow students: “I see so many planning their days so and so. But these are the people going sour and being broken down.” How much this is a reflection of a creative and improvisational approach to practising, or a reflection of the romantic attitude of “rejection of the analytical reason in favour of intuition” (Stromberg, 1990, p.42) is hard to tell. An attitude like this is both theoretically and practically interesting, while it points to the beliefs students harbour about playing and practising. It is, accordingly, also important to include research on their theories or beliefs about practising, their “practice theory”, in future research.

References


THE MICRO STRUCTURE OF PRACTICE

Siw Graabræk Nielsen
Dr.polit-student and Research fellow at The Norwegian State Academy of Music

Introduction

Contemporary research in areas other than the musical domain has yielded the important finding that successful students are strategic (Pressley, Borkowski & Schneider, 1987). They actively engage in strategy use and spontaneously invent increasingly advanced strategies to improve their performance in various subjects (Bråten, 1996). Moreover, contemporary research holds that strategic processing is embedded within a functional system where multiple cognitive and noncognitive components interact (e.g. Borkowski & Muthukrishna, 1992). When this componential system is fully integrated and functions properly, that is, when strategies, metacognition, and domain-specific knowledge are actively coordinated to attain goals, motivated in part by longstanding success in gaining knowledge through strategy use, skilled self-regulation of learning may be said to occur (Borkowski, Carr & Pressley, 1987;Pressley et al., 1987;Zimmerman, 1989). Although there is research focusing on the way musicians master musical works for performance, focusing on individual diversity as it is influenced by learner characteristics, task characteristics, and learning strategies (Chaffin & Imreh, 1994;Gruson, 1981;Hallam, 1992;Miklaszewski, 1989, 1990;Wicinski, 1950), research considering the musician's self-regulation of learning strategies during practice is almost non-existent.

Focusing on the musician's self-regulation of learning and breaking down tasks into component processes and strategies, may contribute to the understanding of how musicians' learn and think during practice. This may lead to improved teaching and assessment of learning and problem solving during practice. Considering the foregoing, the present study investigated how musicians' self-regulates learning strategies during practice as a musical work is mastered for performance. Given that strategic processing has been seen as important in cognitive problem solving (Mayer, 1994;Van Lehn, 1989), this study focuses on self-regulation of learning strategies during practice, which is viewed as cognitive problem solving.

The precise meaning of strategies is an issue still under heated debate. According to Schneider and Weinert (1990), recent research conceives strategies as goal-directed or purposeful processes, originally consciously applied, but normally undergoing automatization as a result of development and practice. In this paper the terms "strategy" and "learning strategy" are used in accordance with the definition given by Weinstein and Mayer (1986) in that they are "...behaviors and thoughts that a learner engages in during learning and that are intended to influence the learner's encoding process. Thus the goal of any learning strategy may be to affect the learner's motivational or affective state, or the way in which the learner selects, acquires, organizes, or integrates new knowledge." (p.315).

The regulatory processes that guides the musician's strategy use during practice are assumed to be activated when the problem being solved is of moderate difficulty (Flavell, 1987;Kluwe, 1987;Sternberg, 1986). In such situations no obvious method of solution is available to the problem solver, and it is important to make correct decisions about the
solution procedure. Mayer (1994) describes this situation as solving ill-defined problems as opposed to solving well defined problems, where both the goal state and the solution procedure is available to the problem solver (in ill-defined problems either the goal state or the solution procedure, or both are unavailable). It seems reasonable to assume that problems during practice are likely to be both well defined and ill-defined. The more specific problems of a particular piece could be defined as ill-defined problems with a well defined goal state, and therefore be assumed to activate regulatory processes under practice conditions (Ericsson, Krampe & Tesch-Römer, 1993; Lehmann, in press; Sloboda, 1985).

By focusing on what problems the investigated musicians' reported and, and what they are doing to solve these problems, this study is concerned with exploring the process of self-regulation of learning strategies during practice and as such the micro structure of practice.

Method

The musicians' investigated were two volunteer church organ students (one male, one female), 21 years old, in their third year at the Norwegian State Academy of Music. Their teacher described the students as gifted, possessing a high level of technical skills at the organ.

The musical works practised were the Prélude from "Prélude et fugue" in B (opus 7) by Marcel Dupré for church organ (student A), and the Salve Regina-movement from the 2nd Symphonie (opus 13) by Charles-Marie Widor for church organ (student B). Both Dupré and Widor is significant composers in the organ literature, and the pieces represent important works from the French romantic period. Stylistically the Prélude by Dupré is written in a very "pianistic" way with fast toccata passages (prescribed tempo by the composer is \( M=112 \)), and with fragments of varying textural complexity. It is notated in 103 measures, and lasted about 3 minutes in the prescribed tempo in the student A's concert performance. The Salve Regina-movement by Widor has both fast toccata passages (prescribed tempo is Allegro) and slower choral passages (prescribed tempo is Tranquillamente assai) with varying textural complexity. It is notated in 78 measures, and lasted about 5,5 minutes at the student B's concert performance. The respectively pieces was well known by the students from other performances and recordings, although no special auditory or analytic preparation took place before the first video-recorded practice session. The pieces formed a part of the students' preparations for the final examinations that were part of their degree programs. However, it must be emphasized that the pieces were selected by the respective student and his or her teacher, and assumed to represent problems of moderate difficulty for the students.

The results presented are based on information gathered in two phases of each student's work with a new piece. The first involving the students' first practising session with a new piece. The second phase took place about a month later when the students were going to prepare it for the final performance. Before this phase each student had already performed the whole pieces for his or her teacher, but afterwards put them to one side practising other pieces for some time. The results presented are based on information gathered during and immediately after four practice sessions (two from each phase) of each student (each practice session lasting about one hour) which was recorded in each student's usual practice rooms, using the instruments they were familiar with.

The techniques for gathering information were adopted from comparable research on learning
and problem solving in fields such as reading, mathematics, physics and second language learning (e.g. Garner & Alexander, 1982; Marfo & Ryan, 1990; McDaniel & Kearney, 1984; Olshavsky, 1976-77; Peterson, Swing, Braverman & Buss, 1982; Siegler & Campbell, 1989). First, in these fields, researchers have developed different kinds of verbal reports (here referred to as concurrent and retrospective reports) that are valuable for gathering information about the thoughts that occur during problem solving (Ericsson & Simon, 1993). In the course of pilot studies these techniques were adjusted to fit the purpose of this study and the naturalistic practice situation. The information was gathered through the use of concurrent verbal reports during practice and retrospective debriefing reports given after practising. These procedures for gathering information were performed according to the recommendations by Ericsson and Simon (1993) and Taylor and Dionne (1994), and they included the conducting of a training session and prompting.

Second, the use of non-verbal methods are seen as valuable. In the present study the students' performances were observed since learning strategies in this context were conceived as originally consciously applied, but normally undergoing automatization as a result of development and practice. Thus, the overt performances of the students was seen as indicating the students' learning strategies. All sequences were recorded on video.

The verbalizations from both the VRDuring- and RRAfter- sessions were transcribed verbatim. According to the theoretical framework, a coding grid was developed for the verbal reports from the VRDuring- session, identifying the following categories: problem recognition, evaluation of performance, and choice of strategies. To determine the degree of reliability of this coding grid, it was applied to all the verbalizations by student A during VRDuring by two independent coders. On the basis of an overall level of agreement between coders of.94, the coding grid was judged to be reliable. Verbalizations from RRAfter and VRDuring were coordinated revealing patterns of self-regulation.

The analysis of the students' performances was based on a study by Kluwe (1987) where the regulatory processes during learning is defined into six regulatory activities concerning the regulation of learning strategies: increasing or decreasing the relative frequency of one strategy applied to a specific segment of the material, increasing or decreasing the average time allotted to the execution of one strategy, and adding new strategies or skipping strategies facing changes in task demands. The performances were registered on the basis of a detailed observational scale.

Results

When information about the students' problem recognition, self-evaluation, use of strategy, and regulatory activities was considered, it became clear that both students were flexible in their strategy use across the phases considering different task aspects in learning the musical material. Both, the characteristics of the musical material and the students' self-evaluation of their performance could be identified as factors of importance for their regulations of strategy use. Mainly the students' regulations of learning strategies could be seen as "trade-off"-decisions whether to emphasize speed or accuracy, or wholes or parts.

The students statements revealed three differing groups of task aspects. Statements concerning the reliability of emerging technical plans and their execution were dominant in the first phase for both students. However, the students' statements show that they were also concerned with other aspects of the material in the first phase. Some of their statements
referred to problems regarding the acquisition of an internal representation of the music. Lastly, there was statements concerning the more expressive qualities of the material to be learned (only student B). However, the relation between the number of statements referring either to technical or expressive qualities of the execution of the material, were changing across the phases. In the first phase, only a few of the students' statements were concerned with expressive aspects. This implies that the expressive qualities of the material either did not represent any problems during rehearsal, or that the students did not yet focus on these qualities in this phase. In the second phase a larger number of statements applied to this aspect, even though statements made by student A still was few. In contrast, in the second phase about half of the statements stated by student B referred to expressive qualities of the material. The material indicate that the task aspects considered as relevant by the students varied across the phases, but that three groups of aspects could be identified in both phases.

Both characteristics of the material and the student's continuous self-evaluation of his or her skills while performing the material led to the regulation of strategies. The material indicate that the students seemed to consider differing features of the material. During the first phase, student A continually made comments on and focused upon segments with "new" elements (that is, "new" in relation to previous patterns processed in this context), while he in the second phase referred to textural changes in the material, and focused on segments comprising a more complex structure. Student B was referring to the texture of the material in the first phase. In the second phase she focused on parts of the piece where the material should be performed in a rather fast tempo, or parts holding a notation key which the student had problems with reading skillfully. This implies that both students considered the characteristics of the material across the phases, even though the features differed between students and between phases. The students' continuous self-evaluation of their performances relied partly on some criteria that the students had defined in accordance with their idea of the material (the outcome), and relied also on some criteria corresponding to what they found as reasonable to expect of their performances in the different phases towards mastery of the piece.

The students' regulation of learning strategies across the phases revealed that the students made procedural considerations involving different different "trade-offs". First, the students seemed to consider whether to practice the material in wholes or parts. That is, the students made decisions that apply to the division of the material into varying segments, or to process larger parts weighing both dimensions in relation to each other. Further, the students seemed to consider whether put emphasis on speed or accuracy in their execution of the material. That is, the students made decisions that apply to a rapid execution on the cost of the accuracy, or to a precise execution on the cost of the speed, and as such weighing both dimensions in relation to each other.

The students differed in their "trade-off"- decisions. In the second phase Student B put equally emphasis on both dimension in each dyad. For instance, the student stressed both speed and accuracy in her execution of the piece, and both divided the material into smaller segments and tried to perform larger parts. In the first phase, especially considering the "trade-off" between speed and accuracy, student B only stressed the dimension of precision and not the dimension of rapidity. Concerning the regulations of student A, he was focusing on one dimension in each dyad in the second phase. For instance, he stressed speed but not accuracy in his execution of the material, and was trying to perform larger parts of the piece in its entirety. This is in contrast to the student's regulations in the first phase where both
dimensions in each dyad were equally emphasized. This implies that the students' regulations of their processing of the material both reveal similar features (as exemplified through the same dimensions in their "trade-offs"), but simultaneously making different "trade-off"-decisions across the phases attending to differing aspects of the final execution of the material.

Conclusion

The purpose of this multiple-case study was to investigate how musicians self-regulate learning strategies during practice as a musical work is mastered for performance. The results indicate that both students were flexible in their strategy use across the phases considering different task aspects in learning the musical material (e.g. technical or expressive aspects of the material, and acquisition of an internal representation of the music). Mainly the students' regulations of learning strategies could be seen as "trade-off"-decisions whether to emphasize wholes or parts, or speed or accuracy. This seems to be in accordance with assumptions made by Singer (1978) in relation to motor skills and learning strategies in sports. Further, both the characteristics of the musical material and the students' self-evaluation of their performance could be identified as factors of importance for their regulations of strategy use. This indicate that these factors are features of the theories of learning strategies that the investigated students held as a basis for their self-regulated learning during practice.

References


VERBAL PROTOCOL ANALYSIS AND RESEARCH ON INSTRUMENTAL MUSIC PRACTICE

Siw Graabræk Nielsen
Dr. polit-student and research fellow at The Norwegian State Academy of Music

Introduction

Recently, the use of verbal protocol analysis has been applied to the study of the cognitive processes involved in musical listening, composing, performing, and conducting (Richardson & Whitaker, 1996). This paper addresses a methodological issue regarding the adaption of this technique for gathering information in research on instrumental practice exemplified through a case study of a church organ student practising a musical work for performance.

The case study investigated how a musician self-regulates learning strategies during practice as a musical work is mastered for performance (Nielsen, 1997). Given that strategic processing has been seen as important in cognitive problem solving (Mayer, 1994; VanLehn, 1989), this study focuses on self-regulation of learning strategies during practice, which is viewed as cognitive problem solving. Although, there is research focusing on the way musicians master musical works for performance, focusing on individual diversity as it is influenced by factors as learner characteristics, task characteristics, and learning strategies (Chaffin & Imreh, 1996; Gruson, 1981; Hallam, 1992; Miklaszewski, 1989), research considering the musician's self-regulation of learning strategies during practice is almost non-existent. In this case study the term "learning strategy" were used in accordance with the definition given by Weinstein and Mayer (1986) in that it is: "...behaviors and thoughts that a learner engages in during learning and that are intended to influence the learner's encoding process. Thus, the goal of any learning strategy may be to affect the learner's motivational or affective state, or the way in which the learner selects, acquires, organizes, or integrates new knowledge." (p.315).

A verbal report is a problem solver's account of her or his own cognitive activities (Ericsson & Simon, 1993). Participants are asked to engage in a task, and think aloud as they solve it, telling whatever is going on in their heads while their work. Ericsson and Simon (1993) differentiate between two types of verbal protocols: concurrent and retrospective. Concurrent verbal reports involve reporting thinking activity aloud as it occurs during engaging in a problem. Retrospective reports are accounts of the actions and thoughts remembered from the problem solving activity, verbalized immediately following this activity. The original verbalizations of the subject are taped and transcribed verbatim. Transcriptions of the tapes result in a protocol that provides the data for process analysis.

Despite the wide-spread use of verbal report data in problem solving research, the validity and reliability of such data have been questioned (Praetorius & Duncan, 1988). There have at a theoretical level been discussions of whether, or under what conditions, cognitive processing can be reliably accessed and reported, and on a practical level, on the use of verbal report data (Taylor & Dionne, 1994). In problem solving research the investigator's control over the situation is severely limited (i.e. the control over the subject's choice of strategies and
action sequences), and the investigator's specification of the problem must take account of how the subject conceives the situation and task in question (Praetorius & Duncan, 1988). There is however, within an information processing framework growing theoretical and practical support for the carefully specified use of verbal report data in problem solving research that offers a strong methodology. That is, that ensures the collection of valid and reliable data, that is consistent with the theoretical framework and research objectives of the study, and that defends any anticipated challenges to validity and reliability (Ericsson & Simon, 1993). The challenges of validity and reliability are referring to conditions and constraints that must be met during verbalization in the problem solving situation and interpretation of the protocol data.

In the following I will focus on how these conditions and constraints relating to verbalization in the naturalistic situation of practising and interpretation of the protocol data were met in order to generate rich and valid data, and describe how the techniques of verbal protocol analysis were adjusted to fit the purpose of the case study presented.

Adjusting to the situation of practising

According to Ericsson and Simon (1993) one of the main challenges of validity is that giving verbalizations concurrently with the cognitive activities, or even knowing that one is to give retrospective reports after the problem solving changes the performance and hence the cognitive processes studied (the effect of verbalization). A second challenge is that the subject may fail to verbalize a considerable part of the information that passes through her short-term memory, or that she uses in the task she is performing (incompleteness). A third challenge is that the verbalizations may report an activity that occurs in parallel with, but independent of, the actual thought process, hence provides no reliable information about the latter (irrelevance).

What specific conditions and constraints had to be met relating to the student's verbalizations of learning strategies during practice? In the naturalistic situation of practising the student is performing ongoing skilled activities, and real-time performance is an integral part of the skill. First, the setting and task design had to administered in such a way that it did not change the student's actual problem solving behaviour. Second, the student's verbalizations had to be structured in such a way that it didn't distort the student's thoughts as learning strategies related to the real-time performance during practising.

It is well known that although we may be perfectly well aware of what we are doing or thinking, we may not always be aware of how we do what we do, or why we think or do as we do (Schneider & Weinert, 1990). Awareness of how we think will typically occur spontaneously only in situations when our otherwise smooth and well-formed actions do not lead to the results or goals desired. According to Flavell (1987), these situations occur when the problem being solved is of moderate difficulty. Considering the foregoing, the task at hand (the musical work) was selected by the student and his teacher, and assumed to represent problems of moderate difficulty for the student. The piece formed a part of the student's preparations for the final examinations that are part of his degree program, and as such the task was familiar to the student, and did not represent an artificial task. The setting
was also familiar in that the student practised in his usual practice room containing a instrument he was familiar with.

Generally, the student was practising without verbalizing his thoughts during practising. In this present study it was important to find a solution that didn't rearrange the naturalistic situation of practising, and that, supposedly, didn't change the student's thoughts during practising. Several solutions were suggested. Based on a study by Olshavsky (1976-77) that focused on verbal reports during reading, the use of physical signals (red dots) following each phrase in the musical material to indicate points at which the student reported the heeded information, were considered. However, it was questioned if this solution would secure the continuity of verbalizations considering the possibility for the student to focus on parts of the piece between phrases, and as such the possibility of long intervals without any verbalizations. Then it was suggested to use sound signals interrupting the student at unpredictable moments in time, at which point they reported the information they were heeding. However, this was considered to be in conflict with the emphasis on a naturalistic situation of practising where sound is a "natural" requirement. It was also suggested to divide the practising situation into smaller units, but this was also considered to be in conflict with the emphasis on a naturalistic situation of practising. The final solution was to use the small pauses occurring during the student's practice session whenever the student repeated some units of the material such as a section, a measure or even a smaller unit, at which point the student reported the information he were heeding. This solution was considered not to change the student's thoughts as learning strategies related to the real-time performance during practice, and it could secure the continuity of the student's verbalizations as learning strategies. This solution, called Concurrent Verbal Reports During Practising (VRDuring), was tested and adjusted in the course of three pilot studies. In addition, some practical guidelines assumed to offer some help in order to maximize the validity of the student's verbalization in the VRDuring-session, were followed. Consistent with the advice given by Ericsson and Simon (1993), the student was instructed to focus on those processes involved in his problem solving during practice and continuously give reports of them as answering the following questions: "What am I thinking?" and "What am I focusing on?", and a training session was conducted. Further, a demonstration video was developed showing the researcher giving verbal reports during practise. During the VRDuring-session reminders to keep verbalizing was given by the researcher.

Referring to retrospective reports, the primary threats to the validity are similar to that of concurrent protocols, but they are heightened considering that the retrospective reports are accounts of the actions and thoughts remembered from a cognitive activity, verbalized following that activity. This is because less of the information heeded in processing is available in long-term memory, and there exits the possibility that the report of a specific episode may be embellished by inference from past processing experiences or be rationalized by the problem solver (Taylor & Dionne, 1994). Ericsson and Simon (1993) suggest that retrospective reports should be elicited as soon as possible following the cognitive activity to be reported. In the present study the main problem was that the retrospective reports had to be reported following the practice session. Considering the maximum period of time between the cognitive activities and the retrospective reports (a practice session lasting approximately about one hour), this was assumed to be a serious threat to the validity of the retrospective reports given by the student. Besides performing the retrospective reports immediately after
finishing the VRDuring-session, it was suggested to offer a stimulus to help the student recall his original problem solving activities. The stimulus utilized in the present study was a video recording from the VRDuring-session showing both the student's concurrent verbal reports and practice behavior. The student watched the video recording during his retrospective reporting. The video recording could be stopped if the student needed more time. Based on a study by Taylor and Dionne (1994) that found that the complementary use of concurrent verbal protocols and retrospective debriefing reports provided frequent opportunities to verify the data reported by problem solvers and to enhance validity in the interpretation of the data collected, reports was facilitated by questions from the researcher. Thus, the cues offered from this video recording also functioned as cues for the structuring of the researchers questions during this session. This solution, called Retrospective Debriefing Reports After Practice (RRAfter), was tested and adjusted in the course of three pilot studies.

Referring to the complementary use of concurrent verbal protocols and retrospective debriefing reports providing frequent opportunities to verify the data reported by problem solvers and to enhance validity in the interpretation of the data collected, the RRAfter-session was also recorded on video, including the video recording from the VRDuring-session, making it possible to coordinate the student's verbalizations from VRDuring and RRAfter. In addition, the student's performance during the VRDuring-session was observed to provide an opportunity to check the reliability of the student's verbal reports as learning strategies.

The verbalizations from both the VRDuring- and RRAfter- sessions were transcribed verbatim. According to the theoretical framework, a coding grid was developed for the verbal reports from the VRDuring-session, identifying major patterns of the student's regulation of learning during practice. The categories identified were: problem recognition, evaluation of performance, and choice of strategies. These broad categories were elaborated by definitions and prototypical examples (see Table 1).

To determine the degree of reliability of this coding grid, it was applied to all the verbalizations during VRDuring by two independent coders. On the basis of an overall level of agreement between coders of .94, the coding grid was judged to be reliable. The coded verbalizations from VRDuring were organized in two different ways. First, to observe emerging patterns of self-regulation, the verbalizations were organized into a scheme showing both the verbalizations and the following performance in succession (see Table 2). Second, in order to provide opportunities for clarification and elaboration of the patterns of self-regulation contained in the information from VRDuring, the verbalizations from VRDuring were aligned with the verbalizations from the RRAfter. One of the advantages of comparing the information in VRDuring and RRAfter segments, was that the RRAfter report frequently made explicit the information implicit in the VRDuring report. The opportunity to discuss the problem solving process with the problem solver was also used to validate the researcher's inferences about VRDuring and RRAfter information.
### TABLE 1
The coding grid

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>DEFINITION AND PROTOTYPICAL EXAMPLES</th>
</tr>
</thead>
</table>
| **Problem recognition**  | the student expresses that he has a problem or the nature of the problem. Included are also statements where the student specifies which features of his performance he wants to change.  
Prototypical examples:  
"Oh, here we come to that tricky bit."  
"That change in hand-position is a bit shaky."  
"Mind the right hand's swift change of position." |
| **Evaluation of performance** | a general evaluation of the performance that doesn't contain any information beyond the student expressing his satisfaction or dissatisfaction.  
Prototypical examples:  
"Now I played an error."  
"That was OK."  
"This was a bit too fast." |
| **Choice of strategies**  | the student documents the action taken during practice and possibly the basis for it.  
Prototypical examples:  
"I'll try from the beginning and see how it goes."  
"Play it through a couple of times so as to get it right."  
"There I'll have to use the third finger." |

### TABLE 2
Organizing scheme for VRDuring-protocol

<table>
<thead>
<tr>
<th>PROBLEM RECOGNITION</th>
<th>EVALUATION OF PERFORMANCE</th>
<th>CHOICE OF STRATEGIES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Verbal instruction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1. I'll try from the beginning and see how it goes.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Oh, here we come to that tricky bit(5).</td>
<td></td>
<td>3. Play it through a couple of times alone so as to get it right.</td>
</tr>
</tbody>
</table>
Conclusion

The concern of this paper was a methodological one. It addresses the adaption of the use of verbal protocol analysis in research on instrumental practice exemplified through a case study that investigated how a church organ student self-regulates learning strategies during practice as a musical work is mastered for performance. The study reported in this paper demonstrated that the techniques for gathering information conducted as Concurrent Verbal Reports During Practising (VRDuring) and Retrospective Debriefing Reports After Practice (RRAfter) are valuable for gathering information about the musicians' thoughts that occur during problem solving in the context of practice.

References:


Singers learning and memorising a new song: an observational study

Jane Ginsborg

(Unit for the Study of Musical Skill and Development, Keele University, U.K.)

Introduction

How do singers memorise and perform songs from memory? A variety of techniques is recommended by singing teachers and writers of textbooks for singers but there is little agreement on effective memorising strategies. Until now, no interview, observational or empirical data has existed relating, specifically, to singers. A phenomenographic study, using interviews and self-reports, was carried out by Hallam (1994) to explore instrumental musicians’ use of different types of memory; observational studies of pianists’ practising and learning behaviours have been undertaken by Gruson (1988), Miklaszewski (1989) and Chaffin & Imreh (1994). The observational study described below forms part of a larger investigation into the interaction of words and music in memory. The primary purpose of the study was to establish an objective and reliable body of data about song memorisation on which to base further investigations. Questions and hypotheses, addressing the effect of singers’ expertise on changes in memorising strategies over time, were derived from three sources: the findings of previous research relating to instrumental musicians, statements made by expert singers in a series of interviews carried out prior to the observational study and a pilot study. The correspondence between musical and practice units, the lengths of attempts, the nature of the strategies used by singers and the errors they made were investigated; some of the findings are discussed below.

Method

13 female singers, aged 19-55 (four music students taking singing lessons, four amateur singers and five experienced professional freelance singers/teachers), took part in the study. Participants were asked to learn and memorise an unfamiliar song, She’s somewhere in the sunlight strong by Seymour Barab (copyright Boosey & Hawkes, 1955), in six 15-minute practice sessions over the course of a two-week period; they were asked to provide a spoken ‘running commentary’ on their activities as they practised, and to record their practice sessions on a 90-minute audio-cassette tape which was subsequently transcribed. Pre-task questionnaires and brief practice diaries were also completed by participants, and post-task interviews carried out. Transcriptions were made of the verbal commentaries, all verbal utterances whether sung or spoken and all musical material whether sung or played. Attempts were defined as melodic phrases or part-phrases; they were coded according to ‘mode’ (see below). Attempts were made as the participants looked at the score (assisted) or from memory. Errors (word/music; assisted/memory) were analysed.

Results

Only nine of the 13 participants carried out six practice sessions. For the purposes of statistical analysis, data from the two participants who carried out four sessions was excluded, as was data from the sixth session. Thus comparisons were made between
individuals and groups on the basis of the performance of 11 participants (student: \( n = 3 \), amateur: \( n = 4 \), expert: \( n = 4 \)) over five sessions.

1. **Correspondence between musical and practice units:** Although the interview study provided little evidence that singers carry out substantial musical analysis, Miklaszewski and Chaffin & Imreh suggest that the music to be practised is divided into sections according to its compositional structure. It was predicted, therefore, that participants would begin their attempts at the start of musical phrases rather than mid-phrase. A one-way ANOVA showed a significant effect of starting point: 84.4% of attempts began at the start of a phrase (\( F (1,8) = 247.7, p < 0.0001 \)). Most attempts began at the start of the song. The second highest percentage of attempts began at the start of the last phrase of the song. An analysis of error frequency per beat suggested that participants’ starting points related to the phrases they found most difficult, as evidenced by the number of errors they made, but the correlation between phrase-starts and errors was not significant.

2. **Length of attempts:** Miklaszewski found that “the more the piece is practised the longer are the fragments selected for attention” (p. 107); Gruson found that expert musicians were more likely than novices to start with longer attempts and focus on details as work progressed. Thus it was predicted only that attempt lengths would change over sessions and there would be differences in attempt lengths between the three groups. There was a main effect of session \( (F (4,32) = 3.07, p < 0.05) \) such that attempt lengths increased from a mean of around four bars (equivalent to one phrase) to a mean of around seven bars. However experts’ attempt lengths did not differ significantly from those of less-experienced participants.

3. **Modes of attempt:** The participants practised the song in eight modes: singing the words, playing the melody, shadowing (singing or speaking and playing the melody simultaneously), counting the beats aloud, singing or playing the accompaniment, speaking the words, vocalising and singing or speaking from memory. These modes can be regarded as strategies, in that the participants chose to use them. The experts in the interview study described learning words and music separately in the early stages of memorising a song. It was therefore predicted that strategies other than singing the words (playing, vocalising, counting, accompanying, speaking) would be favoured, at least by experts, in early sessions. Thus, as use of such strategies decreased as the song became more familiar, so singing the words would increase over time. However, this was not the case. Playing and shadowing decreased for all participants but there was no significant change over sessions in the number of attempts in which participants sang the words, vocalised, counted or accompanied themselves.

There were significant group differences, however, in the use of three strategies. Expert interviewees stressed the importance of metre, as a context for recalling the rhythm of the music, to memory for songs. It was therefore predicted that experts would count aloud more than the other groups. This was upheld \( (F (2,8) = 9.41, p < 0.025) \): experts counted aloud more, in 19.3% of attempts, than amateurs (5.2%) or students (1.6%). As expert interviewees reported memorising the texts of songs as if they were poems, it was
predicted that experts would speak the words aloud more than the other groups. This too was upheld \((F(2,8) = 7.59, p < 0.025):\) experts spoke the words aloud more, in 20.2\% of attempts, than amateurs (5.2\%) and students (1\%). Expert interviewees described preparing songs for performance with the deliberate intention of memorising, rather than working on songs and finding after a time that they had been memorised without conscious awareness. It was predicted, therefore, that experts would make more attempts from memory and that they would begin the task of memorising at an earlier stage than the other groups. Both predictions were upheld \((F(2,8) = 7.96, p < 0.025).\) Experts made more attempts from memory (31.5\%) than amateurs (12.3\%) and students (11.1\%). Mean percentages of attempts made from memory increased from 6\% in Session 1 to 31.9\% in Session 5 (see Fig. 1).

Fig. 1: Mean \% of attempts from memory

A subsequent analysis was made of the number of beats on which errors occurred (assisted/memory, word/music), expressed as a percentage of the total number of beats performed, relative to the total number of beats correctly performed. A comparison was therefore made between the number of beats performed from memory by each group expressed as a percentage of the total number of beats performed. A one-way ANOVA yielded a significant main effect of group \((F(2,8) = 7.2 (p < 0.05)):\) the mean percentages of beats sung or spoken from memory were 38.2\% (experts), 19.2\% (students) and 15.6\% (amateurs).

4 Number of strategies used: Hallam found that expert instrumentalists reported combining a number of approaches to memorisation, while less experienced players relied on fewer strategies. It was therefore predicted that experts would use more strategies than the other groups. This was upheld \((F(2,8) = 9.14, p < 0.025).\) Experts used a mean of 6 modes, amateurs 5.2 and students 4.1. There were no significant sessional changes. It was also predicted that experts would focus on the words separately from the music, on the musical context of the words and melody, and on the task of memorisation itself. Insofar as experts spoke the words aloud more, counted more, and made a greater proportion of attempts from memory, these predictions were upheld \((F(2,8) = 9.14, p < 0.025).\)
5 Percentage of attempts made over all sessions in each mode: A two-way ANOVA yielded a significant main effect of mode \((F(7,56) = 26.1, p < 0.0001)\): means of 64.6% attempts were sung, 42.1% played, 24.27% shadowed, 19% sung from memory, 11.8% accompanied, 9.5% spoken aloud, 9.3% counted aloud and 7.6% vocalised. There was no significant effect of group, but there was a significant interaction between group and mode \((F(14,56) = 2.15, p < 0.05)\) such that students shadowed and vocalised in more attempts than the other groups, amateurs sang and played in slightly more attempts and, as we have seen, experts spoke, counted, sang from memory and, although the difference was not statistically significant, accompanied in more attempts than the others (see Fig. 2).

6 Error data
6.1 Word errors and music errors in all attempts (assisted and memory conflated): Which types of error are most prevalent for each group at different stages of the learning process? Comparisons were made between the percentages of attempts containing one or more errors in each category (word meaning, sound, underlay, omission; music pitch, rhythm, duration, omission) made by each group in each session. There were no significant differences between the groups or over sessions for any type of error. Thus experts appear to be no less likely to make errors of rhythm or pitch, for example, than less-experienced singers. (This is rather surprising, as one might expect professional singers to be more accurate musicians than students and amateurs.) The mean percentage of attempts containing music errors was nearly four times as high as that containing word errors (30.8% music, 8.8% words); on the other hand, participants omitted the words in exactly three times as many attempts as they omitted the music: 1.5% of attempts contained word omissions, while 0.5% of attempts contained music omissions. This analysis does not differentiate between assisted attempts and attempts from memory but presumably errors of omission were made in the latter. It would appear, therefore, that participants found it harder to learn the music than the words but, as they made more errors of word omission than music omission they found it harder to recall the words than the music.
6.2 Word/music errors in assisted/memory attempts (analysed separately): Two-way ANOVAs were carried out in order to examine the main effects of group and session, and any interaction between them, on word errors and music errors in assisted attempts, and, separately, word errors and music errors in attempts from memory, expressed as a percentage of the total length of attempts, calculated in beats. There were no significant differences between the groups or over sessions. Grand means are shown in Fig. 3.

When singers are working from the musical score, then, they are less likely to sing the music correctly than the words: for every five ‘wrong notes’ in assisted attempts participants only sang one beat to a ‘wrong word’. Conversely, when they sing from memory, they are more likely to make mistakes in the words than in the music: participants sang five beats with ‘wrong words’ from memory for every four ‘wrong notes’ from memory. This finding supports the suggestion that music is harder to learn but words are harder to recall.

Discussion
The study was designed to examine differences between expert, amateur and student groups as they carried out the task of learning and memorising an unfamiliar song over the course of five fifteen-minute sessions. Since only one participant in the study was able to sing the whole song accurately from memory by the end of the fifth session, no conclusions can be drawn from the results of the study relating to performance from memory. On the whole, the findings support those reported in the literature. Like the instrumentalists in Hallam’s interview study, the expert singers in the present study combined strategies involving ‘cognitive analysis’ (noting salient features of the music) with ‘automated processes’ (repeating the music to be learned and memorising it largely unconsciously). The singers in the interview study reported that they carry out only the most basic analysis of compositional structure when they set out to memorise a song. This was borne out by the ‘running commentaries’ and data from the post-task interviews of most of the individual participants in the observational study. (The experts, however, made more reference in their commentaries to features of the song such as its harmonic underlay and the similarities and differences between the melodies and rhythms of different phrases; they appeared to use these as cues for recall.) Nevertheless, most attempts made
by the participants (there were no differences between the groups) began on the first beat of a phrase, rather than mid-phrase. To this extent, at least, then, singers observe the phrase structure of a song, made explicit by bars' rest between phrases in the melody and word cues, when they divide the music into practice units. Gruson found that expert pianists increased their use of the strategy of repeating sections as their work on a piece of music progressed, while Miklaszewski observed his subject working on, progressively, fewer but longer sections. In the present study, there was no difference in attempt lengths between groups, but there was a general, if small, increase over the five sessions. Analysis of error types suggested that no group was better or worse than any other, overall, at learning different features of the words or music. However all the participants made many more music errors than word errors. Separate analyses of word and music errors in assisted attempts and attempts from memory, in terms of length in beats, revealed more music errors than word errors in assisted attempts, and more word errors than music errors in attempts from memory. Overall, the findings support the suggestion that music is harder to learn than words, but that words are comparatively harder to recall than music: singers’ fears that they will forget the words in performance, while being able to ‘keep the music going’, are well-founded.

When comparison is made between the responses of the experienced professional singers in the interview study and the actual behaviours of the observational study participants, there is clearly a mismatch between what singers think they do, and what they actually do. For example, all but one of the five singers in the interview study described studying the text in detail before beginning to work on the music. However only one of the 13 participants in the observational study carried out detailed analysis of the text, and this was not done until the third session.

Further investigations will be experimental. The intention is to isolate the different strategies by asking participants to use one at a time; analyses of error frequency and type, in attempts using each strategy, can then be compared.

References


The Significance of Aural Development in the Performing Musician

Alison McNeil
Department of Music, University of Huddersfield
Queensgate, Huddersfield. HD1 3HZ
a.mcneil@pegasus.hud.ac.uk

Abstract

"We teach ear training and we have taught it for some time, but what degrees of perceptual faculty are required to be 'effective' as a musician?" (Carlsen, 1969)

This study is primarily concerned with the Associated Board of The Royal Schools of Music (ABRSM) who have designed syllabuses during this century to develop instrumental performance ability, culminating in an examination of performance achievement. The ABRSM performance examinations incorporate 4 constituent parts; the area concerned with here is a series of subtests that fall under the umbrella of 'aural tests'. As the fundamental factor for assessment is performance, the other constituent parts of the examination are intended specifically to, and should, reflect the skills, developed through training and conceptualisation, which are central to the make-up of a convincing performance.

The aim of this study is to assess whether aural skills, culminating in their testing as part of the ABRSM practical examinations, actively influence, inform and enhance performance skills; in so doing, bridge the ideological and knowledge divide between the applied disciplines of music, psychology, and education between aural ability and performance.

The first part of the study involved a series of open-ended interviews with teachers, educationalists, performers, examiners, adjudicators, and psychologists concerning the skills, abilities and issues involved in an effective musical performance. Subject responses were subject to a content analysis from which data was collected concerning the relevance of aural ability to the training and execution of performance acumen. From the data compiled, a proposal for the redevelopment of the ABRSM aural tests will ensue, focusing upon aural skills that develop and enhance performance ability. Finally, a learning program for the training of these skills will be developed.

Study

As music psychology has become more accepted and respected in the twentieth century, and under the influence of 'comparables' in cognitive psychology such as intelligence tests, a number of standardised tests of musicality have been developed. Seashore (1919) provided the impetus for this work with his assessment of musical talent, followed by numerous other 'constructs' developed to establish the validity of a variety of test functions (Revesz, 1920; Wing, 1939/1961; Gordon, 1965, 1979a, 1984, 1987; Seashore, 1938; Shuter-Dyson and Gabriel, 1981; Bentley, 1960; ABRSM, 1920/1993/1995). It is the later of these, the ABRSM examinations board, that this project is concerned with.

Through the last decade, aural skills have been assumed as being pertinent to a musicians' training and performance acumen (Sloboda, 1986; Elliott, 1987; Clarke, 1988; Salaman, 1994); the assessment of performance ability, musicality and musicianship all incorporate tests to assess aural ability through an oral response to aural stimuli. In particular, aural
Acuity has been, and is, an expected part of a performer's ability, demonstrated through the variety of aural tests present in the assessment of performing musicians. In particular, the ABRSM aural tests have been revised and re-revised, with the expressed intention of placing the emphasis on 'perception and listening with awareness' (ABRSM). Inevitably, these modifications have generated an intense discussion about the role and purpose of aural training and the testing of these skills. Teachers, pupils and educationalists alike greeted these aural syllabus changes with ambivalence: some remain firmly rooted in the traditional camp of aural testing, others finding these new tests somewhat refreshing, relating more towards actual performance skills. My objective is to assess whether aural skills, culminating in their testing as part of the ABRSM practical examinations, actively influence, inform and enhance performance skills; in so doing, bridge the ideological and knowledge gap between the disciplines of music, psychology and education between aural ability and performance.

In order to ascertain the answer to Carlsen's question we first need to establish valid performance criteria for assessment purposes. Open-ended interviews were undertaken with subjects consisting of board members of the ABRSM, instrumental and vocal teacher, active musicians, music psychologists and other educationalists. The interview schedule raised issues concerning the criteria used to assess a good performance, how to train these skills, whether aural skills are necessary for a performance to be effective and if so, how these aural skills are defined and taught. The data collected from the interviews (mean time = 1.5 hours) was subjected to a content analysis.

Results

The results of the content analysis show that performance criteria and the training of these skills relies heavily upon the training of aural skills. Outlined below are some of the key areas cited by participants:

Memory was specified by a significant amount of interviewees (84%) as being imperative for an effective performance, the skill of memorisation being attained through aural development. However, there are many types of memory available to the performing musician. Those found to be significant are kinaesthetic memory and aural memory ('aural imaging' see Pratt, 1990) and in conjunction with this visual memory. The ability to aurally internalise a visual pitch, resulting in accurate externalisation on one's instrument or voice, is a recurrent theme throughout these interviews. Paradoxically, the method teachers employ to develop the ability of aural internalisation is through encouraging performers to sing; singing being the most controversial issue surrounding aural testing where by candidates are expected to sing back responses.

The concept of aural ability relating to performance skills created a tension for interviewees: the majority of subjects being unable to divorce the concept of aural testing from aural skills. Indeed, many regarded performance and skills assessed in aural tests as completely separate; in some cases aural tests were seen as bearing no relationship to performance ability. One wonders whether aural tests have skewed the perception of aural skills rather than enhanced it.

Once interviewees had grasped the distinction between aural testing and aural skills in performance, it became apparent that subjects placed great emphasis upon the necessity to integrate the training of aural skills with performance acumen. The underlying reason given is that performers' need to understand how these two phenomena are connected, and with this knowledge their performance should be enhanced.
Subjects' noted that as a performer's ability to coherently verbalise about musical aspects develops, so too does performance ability. However, the majority of interviewees also reported that pupils already know 'in their heads' what's going on but it's getting to know adequate language to express it'. There is an anomaly here. If pupils know yet cannot express, why should their performance improve as their expressive abilities develop? Could it be the connection of understanding that is made? Many seem to think so. A recurrent theme throughout this study is the belief that an understanding of the link between aural ability and performance skills is crucial to a musician's development. Findings reveal that the ability to understand this link, through language and knowledge is the key area. Primarily, subjects perceived understanding as the ability to attach meaning to what one is doing in performance, negating the technique of rote learning.

The description of understanding as I see it describes what music is for, not necessarily how it works. This expresses the need for a knowledgeable understanding that is gained through experience; conversely the requirements for aural tests as a component of the ABRSM performance examination necessitates the ability to verbalise about music. Indeed, the question as to what verbalising is for merits a somewhat vague answer; why do we need to talk about note names? As one subject retorted 'we learn by listening...not through verbalising'. I propose that it is the conceptual connection of understanding, gained through knowledge acquisition and expressed through language, that is seminal to this. Indeed, throughout this study the belief is held that an understanding of the link between aural ability and performance skills is crucial to a musician's development.

A paramount area of discontent concerning aural testing is that of timbre: timbre effects a performer's ability to respond to aural tests, the tests being administered on the piano with candidates responding to the various tasks with clapping, tapping, verbalisation and singing. In addition, interviewees related timbral problems to specific instruments, often proposing ways in which administering and responding on one's own instrument would benefit the performer. Indeed, many subjects perceive the ability to sing back a response and play back a response as two very different cognitive processes; the question is which process is more akin to performance ability? Both the ABRSM and subjects acknowledge that the piano is used to administer tests due to practical reasons. However this cannot be a valid or reliable reason for using the piano as a gage of performer's aural abilities.

One solution to the timbral problem may be for tests to be administered on the candidates' instrument and for the candidates to respond using their instrument. Indeed, during this study, additional subjects were asked to administer aural tests as an integral part of a pupil's lesson, on the pupil's own instrument. A significant amount of performers showed improved responses, some being able to respond where they had not been able to previously.

Conclusions

Aural ability undoubtedly underlies many performance skills; proposed either consciously or subconsciously by interviewees. Throughout this study an effective means of training aural skills was noticeably lacking. This project continues to explore how performance skills are achieved, what techniques are used to fulfil this, culminating in a more relevant means of assessing aural ability. More effective teaching methodologies will be explored in order to facilitate the development of these skills, enabling a more fruitful learning process for the performing musician; this will require a more tangible link between the concepts of aural ability and performance skills.
The results indicate several issues to pursue in the later part of this research, notably the training methodologies, learning strategies and assessment of aural development in the performing musician. Proposals outlining ways of assessing aural ability within the context of performance, complimented by separate tests that are more relevant to the performer are in progress. In conjunction with this, an examiner's assessment strategy is being developed, focusing upon the skills within performance that denote the possession of aural ability.

The pragmatics of musical training and examination need to be informed by appropriate 'psychology of music' theory and research, instead of doing what people simply assume to be necessary and effective. Combining the skills and experience of both disciplines is imperative if we are to progress.

__Select References__


The Meaning of Music and the Understanding of Teaching and Learning in the Instrumental Lesson

Anna Reid, University of Technology, Sydney.

Teachers and students of instrumental music understand teaching and learning as it relates to music in a variety of qualitatively different ways. This paper will discuss the meaning that instrumental teachers and students ascribe to music within the context of the instrumental lesson and how their conceptions of teaching and learning are expressed as a result. This is part of a larger study on teachers' and students' conceptions of teaching and learning music. Phenomenographic analysis of the interview transcripts with tertiary instrumental teachers and students from an Australian school of music have resulted in the uncovering of conceptions of teaching and learning within this context. That participants not only expressed variety in their understanding of the nature of teaching and learning in music but in the way music itself was understood. In essence the most sophisticated views of music related to the most sophisticated views of teaching and learning. The meaning ascribed to music and to the teaching and learning of music through the teachers' and students' description of their experiences of the instrumental lesson form the basis of this paper.

Students' conceptions of learning were first identified by a study by Säljö (1979). Students described their understanding of the experience of learning in five qualitatively different ways; as an increase in knowledge, memorising, acquisition of facts, developing meaning and understanding reality. A sixth conception was identified by Marton, Dall'Alba and Beaty (1993); learning as changing as a person. In a theoretical analysis Biggs (1990) distinguished parallel conceptions of learning and teaching that related a quantitative conception of learning (learning is a matter of how much is learned) with a quantitative conception of teaching (teaching is the transmission of knowledge). Institutional and qualitative conceptions of teaching and learning were also paralleled.

Theories of teaching in higher education have been expressed in terms of teaching as transmission, as organising student activity and as making learning possible (Ramsden, 1992). Conceptions of teaching in the science and social science have been described as teaching as supporting student learning, teaching as an activity aimed at changing students' conceptions or understanding of the world, teaching as facilitating understanding, teaching as transmission of knowledge (Samuelowicz & Bain, 1992). This research aims to describe instrumental teachers' and students' conceptions of teaching and learning (Reid, 1996) in relation to their understanding of music as it is taught and learned.

Stubley (1992) suggests that music may be understood as a way of knowing where 'the individual, social, and historical dimensions of the musical experience interact with different contexts and situations'. Musical understanding as a series of
'networks' includes the ability of musicians to 'produce, perceive and reflect upon their work' in order to 'think like a musician' (Davidson a & Scripp, 1992). Part of that 'way of knowing' would include relations to music as it is understood as something to be taught or learned. A 'way of knowing' in this study is expressed as the 'Music Object' which is the way notions and experience of music are reconstructed for the purposes of teaching and learning. The Music Object both informs and undergirds the Instrumental Teachers' and Students' Conceptions of Teaching and Learning.

Method

Phenomenographic interviews were used to allow the participants to explore the ways in which they were aware of and gave meaning to their experiences of teaching and learning (Marton, 1981). Phenomenographic interviews are in depth and open ended allowing the participants to fully explore their understanding of the object of research, in this case teaching and learning instrumental music. Each interview was of approximately one hours duration. Transcripts of the interviews were analysed as a group to determine the variety of meaning that was ascribed to teaching and learning across the whole. The qualitatively different conceptions were defined as categories of description that characterised the qualitatively differences relating each category with each other in a structured and logically related hierarchy. This resulted in three sets of categories of description: Instrumental Students' Conceptions of Learning, Instrumental Students' Conceptions of Teaching and Teachers' Conceptions of Teaching and Learning. The sets of categories of description emerged from the transcripts themselves. Iterative reading and re-reading of the transcripts allowed the researcher to uncover the characterising differences in meaning across the transcript group as a whole. Student and teacher groups were treated separately.

The participants in the study came from a tertiary music school in Australia. Ten instrumental teachers involved were asked to select two each of their students who they thought learned in different ways. Teacher participation was voluntary and participants came from several instrumental and style groups. The common context for the research was that all had to be teaching at the school at tertiary level either in a full time or part time capacity. All of the students were full time tertiary students.

Instrumental Students' Conceptions of Learning.

Conception A: Learning an instrument. In this conception the object of learning music is the instrument itself. The students focus on the technical skills required to play the instrument. They rely on their teachers to organise their repertoire and practice schedules. Manuscript is seen as a series of technical problems that have to be individually solved. The outcome of the learning experience is the ability to demonstrate technical skill on the instrument and to pass examinations.

Conception B: Learning an instrument and some musical elements. In this conception the instrument still remains the focus of the learning experience but some musical elements are included. As in conception A the focus is on the acquisition of technical
skill combined with a dependence on their teacher to choose repertoire and organise practice schedules. Manuscript is seen as a series of individual technical problems that must be solved. Once technical solutions have been found musical elements, such as phrasing or dynamics, are added. The outcome of the learning experience is technical proficiency in an exam situation.

**Conception C: Sound.** Musical sound is the focus of this conception. Students reflect on their teacher's advice on technique and stylistic interpretation of music. Technical proficiency on the instrument is seen as a vehicle to enable correct playing of the music. The outcome of the learning experience is to be able to play the music with correct technique and musical style in a performance situation.

**Conception D: Sound and communication.** In this conception music is seen as a means of communication with an audience. Written music has an inherent meaning that is expressed by the student using the instrument as a medium. Technique is seen only as a tool through which musical meaning is expressed. The outcome of their learning experience is to express the implicit musical meaning of a work to an audience.

**Conception E: Sound, communication and personal meaning.** Students expressing this conception see music as a means of self expression and communication. Learning is expressed as an experience that requires the student to reflect on musical knowledge and assimilate musical ideas into a performance that both communicates with the audience and expresses personal meaning. The instrument itself is seen only as a vehicle of self expression which is subject to the greater need to express personal meaning through music. Engagement with music itself is seen as a method of continuous personal development. The outcome of the learning experience is to communicate personal meaning and interpretation of the music to an audience through performance.

**Teachers' Conceptions of Teaching and Learning.**

**Conception One:** Teaching is disseminating the teachers' musical and performance experience. Students learn through being exposed to the teacher. Characteristic of this conception is the belief that it is the musical and professional life of the teacher that the teacher thinks the student wishes to emulate. The role of the teacher is to tell of her experiences and to demonstrate her craft. Students learn by copying the teacher's performance. Music is seen as a collection of technical and musical packages that need to be transferred to the student. Teaching is seen as a formative experience for the student and the outcome of the teaching and learning experience is to provide a background for the student's future development.

**Conception Two:** Teaching is passing on the teacher's experience of music and performance. Students learn by using the teacher's methods and adapting them for their own particular needs. As in Conception One the teachers' performance and musical experience are the focus of the teaching and learning experience. However students are expected to practice according to the teachers' preferences and to try out
the teachers' methods. Learning is considered to be an assimilation of the teachers' ideas and integrating them with the students' ideas. The teachers' ideas are adapted for each student's particular needs and the student learns by copying and practicing the ideas. Music is seen as a collection of ideas that need to be given to the students. The outcome of this teaching/learning experience is student success in assessment.

Conception Three: Teaching is an exchange of experiences and musical ideas with the student. In this conception teaching is seen as recognizing the diversity of student experiences and the teacher's role is to help students build upon their experiences. Students learn by being encouraged to experiment with different solutions to musical problems with the range of ideas being developed and expressed by both teacher and student. Music is seen as a combination of physical, intellectual and artistic skill. The outcome of the teaching/learning experience is for students to show an understanding of and be able to demonstrate the technical qualities of the instrument combined with intellectual and artistic interpretation in performance.

Conception Four: Teaching is an exchange of experiences and musical ideas with the student that changes the way both the teacher and the student think about and engage in music making, communicating and how they see the world. In this conception teachers and students work together to develop new ideas on performance and musical understanding. Music is seen as a "reflection of the world" and the purpose of teaching and learning is to enable both teacher and student to expand and develop their world views. The outcome of the teaching/learning experience is when technique, intellect and musicality are combined to produce a message that communicates with an audience expressing the students' emotions and view of life.

Instrumental Students' Conceptions of Teaching.

Conception i: Teaching is demonstrating musical techniques and the teacher's experience. In this conception students depend on the teachers to provide the lesson content and to direct the activity within the lesson. The transference of technical skills and the teacher's personal experience as a musician are the focus of this conception.

Conception ii: Teaching is expressing musical meaning and the teacher's experience. The students' dependence on the teacher is replaced in Conception ii by a recognition of teacher influence where the teacher's role is to promote student independence in learning and performance.

Conception iii: Teaching is supporting student learning, encouraging independent student expression and preparing for life. The intention of Conception iii is to promote student self expression by assisting the students to develop meaningful relations between the instrument and the music's meaning. Teaching and teachers in this view are influential still but the responsibility within the teaching and learning encounter does not remain the sole domain of the teachers, the students rather play an active role in deciding what, why and how music should be taught.
Discussion

The unique outcome of this study was that the three sets of categories of description were found to be linked together by the participants' shared understanding of what music is as it is either taught or learned. Music, as it is constituted to either teach or learn, is defined as the Music Object. The meaning given to the Music Object plays a pivotal role in determining how students and teachers understand and define learning within the instrumental lesson.

The Music Object in this context does not relate to 'what' is to be taught or learned but is rather how music is defined in order to be taught or learned. The implication of this is that 'music' as it is understood by musicians is redefined when it is related to teaching and learning. The Music Object may be divided into three related components: the technical (either physical or notational) which is an extrinsic aspect, as sound and communication which emphasises extrinsic meanings, and as personal meaning which is intrinsic and related to a view of the world and personal meaning. The Music Object as technique relates to the physical aspects of playing the instrument as well as notational elements such as phrasing, accents or articulation. Sound and communication are related to the belief that each piece of music has an inherent meaning that is constituted by such things as style, period, harmony and composers' intent. It is the music's inherent meaning that is communicated to an audience. Music as personal meaning could involve aspects of the first two components of the Music Object but these aspects were reinterpreted by the participants through the notion that music is a way of expressing personal meaning and understanding of the world through music performance. It is the musicians' ideas that are expressed through the music's inherent meaning.

The Music Object emerged as an artefact of the research, the focus of the research being on the variety of ways that the participants understood and gave meaning to teaching and learning music. Hence the Music Object was inferred by the participants as part of their context for describing their understanding of their experience of instrumental teaching and learning.

The Music Object is a link between all three sets of categories. The technical Music Object can be seen to relate to Instrumental Students' Conceptions of Learning A and B, learning an instrument and learning an instrument with some notational elements; to the Teachers' Conceptions of Teaching and Learning, One that teaching is disseminating the teachers’ experience and that students learn through exposure, and Two that teaching is passing on experience and student adaptation; and ISCOT I where teaching is demonstrating. These five conceptions are the least sophisticated and ignore the possibility that music can be anything but technical. In this regard the participants' understanding of the Music Object is that music as it is taught and learned is only about technical and notational elements. More sophisticated categories where learning is about learning about the music's meaning and communication of that meaning, and where teaching is described by the teachers as an exchange of experiences and by the students as an expression of musical meaning and experience, include characteristics the less sophisticated conceptions as part of their own. These
characteristics are reinterpreted in light of the more sophisticated characteristics and accommodate the notion that music itself has a meaning to be expressed (the second Music Object - musical meaning and communication). Similarly the most sophisticated categories, readily accommodate those that are ‘beneath’ them but understand them and relate them to a view of music that is holistic. For instance students who express their learning in terms of the expression of personal meaning may describe some aspects of learning in terms of technical development but unlike the least sophisticated conceptions technical development is not the intended outcome but a means to the end of communicating their own understanding of the music as it relates to their personal experience.

The participants’ understanding of the experience of teaching and learning instrumental music is related to the meaning that that ascribe to music itself as it is reconstructed to be taught and learned. The link between the depth of the participants’ understanding of the Music Object and their description of their conceptions of teaching and learning is explicit.

Bibliography.


Reading Beethoven: Eye Movements Investigation

Isabelle Servant, Thierry Baccino & Jean-Louis Leleu
University of Nice Sophia-Antipolis

Introduction
Eye movements analysis when reading ordinary printed texts have been the subject of many investigations since the first measurements by Erdmann and Dodge in 1898, and reader’s behaviour is on the whole well understood (Rayner & Pollatsek, 1989). However, there are relatively few studies on how a musician is reading a dual-staved score, and fewer still when reading a fragment of real, artistic work by a great composer such as Beethoven, that is, a natural sight-reading situation for a professional or semi-professional performer. This is quite easily understandable: even in tonal music, rather well apprehended in its system by both musicologists and psychologists, fragments are liable to be exceptionally rich and complex. In that case, experimental procedure might be difficult to use because of the large number of uncontrollable factors, while a reasonably wide skill in musicological analysis is obviously demanded to understand them.

Beethoven’s Bagatelles (op. 33, op. 119, op. 126, in Gustav Henle Verlag, Urtext, München), appear to us as exemplary pieces for such an experiment, in spite of their relative left-hand technical easiness, not quite representative of a XIXth century keyboard language: they are extremely concise and not too difficult to read, and still present no resemblance whatsoever with a traditional and predictable piano-forte drill, such as Czerny’s, for instance. We then hope to make subjects meet with highly significant facts, quite frequent in Bagatelles, even in a very short musical span.

An hypothesis have been made that subjects, through a great number of fixations and/or regressive saccades, would execute in these occasions a heavier cognitive task than normal. To achieve this, we had to establish a second-and parallel-corpus which would present, in the same place as our chosen significant fact, a proposition of a non-ambiguous, less-significant version. Comparisons between the two versions should then give a hint about the manner the composer’s stylistic features would be apprehended by musicians. For instance, text 5 (p 26, n° 7, bar 1-8) shows as a beginning a repetitive left hand formula which would be common-place if it did not last nearly four bars, this length emphasizing the dynamic and suspensive character of the fragment; modified version presents a shorter formula (two bars only); or text 4 (p 24, n° 6, section two, bar 9-13), rather cantabile and predictable in its beginning, contains an ambiguous G note in a bar, where it seems that the text might develop, though temporarily of course, into two different keys, A major or D major. Modified text orients the text to A major, in a non ambiguous and less interesting perfect cadence, D major occurring immediately after.

Method

Subjects:
8 skilled music readers from the Department of Musicology of the University of Nice served as participants.

Procedure:
Subjects were required to sight-read 6 printed scores picked up from Beethoven’s Bagatelles and chosen for their processing difficulties. All scores were dual-staved and two versions were constructed: an original version keeping the richness of Beethoven’s writing using some sort of (tonal) ambiguity and a modified version taking off the ambiguity. During reading, eye movements were monitored by means of an infrared photoelectric system (Bouis), digitized
every millisecond and stored on a computer for later analyses. After each reading, subjects had to answer a question in order to get a focalised attention during sight-reading.

Results
Quantitative analysis

Readers' eye movements were parcelled out into first pass reading and reinspections. First-Pass reading refers to all eye movements made during the initial reading of the score (following a left to right saccade). Reinspections involve all regressive fixations (following a right to left saccade) both within bar line and across them. The number and durations of fixations were analysed according the level of difficulties of scores (original vs modified version) and for each bar-line.

From an oculomotor point of view, a striking difference between scores reading and text reading exists. While in text reading eye movements follow a strict left to right direction using sometimes regressions to retrieve an information in the previous text (10 to 15% of the time) (Rayner & Pollatsek, 1989), the pattern of eye movements during sight-reading shows a constant up-and-down movement between the two staves within which a considerable amount of regressions are made. Around one fixation out of three is a regressive fixation (32% of all fixations) either within the same bar-line or across bar-lines. Fixation durations are comparable with text reading with a mean duration of 243 ms. (See figure 1).

When we look at the oculomotor behaviour between the two staves, musicians used more fixations on the right-hand stave than the left-hand stave \( T(181) = 2.39 \ p<.01 \) showing the 'relative' easiness of the left-hand in these fragments. However, this difference in number of fixations is associated with their duration, fixations on the right-hand stave took longer times than left-hand stave fixations \( T(181) = 2.68 \ p<.01 \).

As far as Beethoven writing is concerned, there are more fixations on the original versions than the modified versions \( T(168) = 2.70 \ p<.01 \) but this difference is mostly due to the increase in the number of regressive fixations involved in reading Beethoven's scores \( t(60) = 2.23 \ p<.01 \). Fixation durations across these two versions were approximately identical. An analysis of eye movements on the precise bar-lines inducing the ambiguity showed the same results suggesting that the difficulty in sight-reading for Beethoven's scores is spread out on the entire line and not only focalised on a predefined region.
Figure 1: Typical eye movements triggered during dual-staved sight reading (Fragment 4.1). The graphic at the bottom indicates fixation duration (on X’axis) as well as saccade amplitude (Y’axis) both on horizontal (a) and vertical (b) eye movements.
Short qualitative description of text 4.1 (see above).
The reading behaviour of this fragment appears rather distinctive of a classical reading
behaviour in a typical sight-reading contest. For the same text, some other subjects were even
more efficient because their extensive examination of the score was preceded by a preliminary
overview of the entire line. The role of this first reading seems to localize on the score where
problematical musical facts (such as inflexions or difficult rhythms) took place in order to re-
examine them a posteriori. Fixations do not begin on the first sign, but are grouped on the
right hand melody in bar 1, then on the entry of the left hand. In bar 2 a rapid analysis of the
perfect cadence in E minor is made in comparison with the bass (A sharp needs two fixations).
The natural A in bar 4 triggers a regressive saccade on the bass and this process will be
repeated to take off the ambiguity on the following G note (sharp or natural?). Indeed this
ambiguity can only be solved by reading the bass perfect cadence in D major, which means the
G will be natural. This difficulty is caused here by Beethoven’s use of a sixth and fourth chord
on beginning of bar 4, which could logically be followed by an A major cadence. After reading
bar 6, the subject makes regressive saccades again on bar 4, showing its importance and
difficulty.
Many regressive saccades are clustered on the signature of the key. This behaviour is largely
used by subjects, and very often causes a quite wide covering of this zone with fixations and
regressive saccades When meeting an inflexion, the subject may check key informations with
regressive saccades. These groupings of fixations do concern the inflexions that define the key
of the piece, and they are not visible in C major, for instance (no inflexion).

Conclusion
This experiment was a first step into the study of sight-reading using eye movements paradigm
in a natural context (real dual-staved score). Oculomotor behaviour during sight-reading
appears very different to text reading. More reinspections need to be made to check the
coherence of the musical structure being elaborate as eyes are progressing on the stave
suggesting a high level control to integrate every visual partial information. One question of
interest of course would be to know what would be the eye-hand span in that context (i.e, the
delay between the fixation and the realisation of the note on keyboard). Further experiments
need to be carry out also to investigate whether eye movements are synchronized with
pulsations or rhythm. From our data, it appears no direct relationship between the spatial
pattern of eye movements and temporal considerations of the score, musician take a maximum
of cues in a minimum of time, in order to reconstruct an a posteriori temporal and as exact as
possible pattern (these results already partially present in Kintzler, 1994, and as a strong
The role of eye movements during music reading

Sophie Furneaux and Michael F Land
Sussex Centre for Neuroscience, Biology Department,
University of Sussex, Brighton BN1 9QG, UK.
S.Furneaux@sussex.ac.uk

Summary:
A head-mounted device was used to study the eye movements of pianists sight-reading dual-staved music. All subjects read the two staves by alternating between them, indicating the presence of a 'buffer' within the system, where information is stored and reassembled until motor execution is required. Also, the eye-hand span (the separation between eye position and hand position) was measured in two ways: using time (the 'time index') and using note units (the 'note index'). The mean time index for a performance is independent of skill, but the standard deviation of the time index is not: within a single performance, novices show much greater variation in their time indexes that skilled musicians. In addition, faster tempos give shorter time indexes, for all subjects. The mean note index is mainly dependent on ability, although the standard deviation is similar for all subjects.

Playing a musical instrument from a printed score poses special problems for the oculo-motor system. The printed material must be fixated, deciphered and processed in a sequential manner, and then reorganised for motor output. This planning of information uptake is unconscious, and must occur within a pre-set time-scale, often very brief, so the timing and coordination of the internal processes are critical for accurate, continuous performance. The process is made even more difficult in piano-playing where there are two staves of visual information to be read and performed simultaneously.

Eye movements provide a 'window' on the systems underlying visual tasks, and as such, serve as excellent tools for studying all types of visual behaviour. Substantial studies on text-reading have provided an abundance of information, both on the nature of eye movement control, and on the cognitive organisation underlying reading (see Rayner & Pollatsek, 1992, for a review). In contrast, very few eye-movement oriented studies on music-reading have been undertaken (see Goolsby, 1989, for a review), and there has been little improvement on the groundbreaking work of Weaver (1943) and Van Nuys and Weaver (1943).

In order to describe the eye movements of pianists, the way that the human ocular system acquires visual information from the world, must first be briefly examined. The role of the visual system is to provide detailed, cohesive, 3-dimensional information about the surrounding world from the 2-dimensional images on the retinas. Eye movements play a important role in this task, because resolution of the finest detail is only possible in the fovea, which subtends approximately 2° of visual angle. Therefore, this part of the eye must be directed towards areas of interest in order for to extract information requiring high acuity. In normal, stationary, scenes this is done via a combination of saccades and fixations. Saccades are rapid, conjugate eye movements, typically lasting less than 50msec, and during which vision is suppressed.
Between each saccadic movement, the eyes are kept stationary with respect to the point of regard, producing fixations. Freely moving individuals normally make three to four saccadic eye movements per second (Rayner, 1978, 1984).

For this study, the eye movements of pianists were recorded via a head-mounted video camera system (Land, 1993). The device is non-intrusive, light in weight and fully portable, and so allows full and normal movement of not just the head, but of the whole body during performance. It records a split-screen image of the subjects' eye and the scene ahead, via two mirrors. A part-silvered mirror is positioned below the camera, which places the virtual position of the camera lens in the left eye orbit. In this way, the forward view recorded by the camera is the same as that seen by a forward-facing eye. The second mirror is concave, and images the left eye into the lower part of the camera field. A computer program that models the iris of the eye is used to obtain the eye fixation direction for each frame of the video (50 Hz). With careful analysis, this process is accurate to approximately 0.5°.

Subjects were pianists of various abilities, and were required to sight-read the presented music. Some experiments involved the imposition of a tempo, in which case a metronome was used: this was silenced after the first full bar of performance.

All subjects were found to read music in a discontinuous and intermittent manner. No simple, temporal relationship was found between the initiation of a saccade and the striking of a note. This indicates that the eye movements themselves do not provide any actual rhythmic or 'beat' information. However, saccades were found to land on, or near to, note heads, showing that note location is seen and used (see figure 1). Also, fixations were longer and more irregular than those found during text reading (mean 400msec, s.d. 300msec), suggesting that music-reading is less of a mechanical process and more affected by ongoing cognitive processing. In agreement with this, fixations were generally longer when the music was melodically, harmonically or rhythmically difficult. This suggests that a new saccade is made only when the information from the previous fixation has been processed and released to the 'buffer' (Kinsler and Carpenter, 1995). Thus, it appears that the timing and destination of each saccade is determined by three things: an overall strategy; by the seen locations of the notes themselves; and by the state of ongoing cognitive processing. This cognitive component must also involve auditory feedback.

All subjects were found to read the two musical staves separately. The path of eye fixations follows a zigzag route, with the gaze alternating between the notes of the upper and lower staves, rather than using centrally located fixations to read both staves simultaneously (see figure 1). This agrees with the findings of Weaver (1943) and Petzold (1995). This result means that the information from each stave is processed separately, to be combined later on. In other words, the two sets of information are obtained individually, in serial form, but are played simultaneously, in parallel. Clearly, some temporal reassembly must occur between input and output, and early 'parts' of the internal system must be duplicated - one for each stave. This finding confirms the necessity for some sort of 'buffer' or 'store' (Kinsler and Carpenter, 1995).

The length of time that it takes to process, store and execute a scored note can be found by determining the delay between the first fixation of that note and its subsequent performance.
This is the 'eye-hand span', measured in time. To produce this 'time index', sound spectrograms of the auditory signal were time-linked with the eye movements made, via an electronic 'clapper-board'. The time index varies considerably throughout the performance of a piece. However, the mean time index of subjects of differing abilities is remarkably similar - usually about 1 second. Enforcing a particularly fast tempo will reduce the mean time index to 0.6-0.8s, whereas an especially slow tempo produces longer time indexes, in the range of 1.2-1.5s. Again, this is true for all subjects, showing that the mean time index is skill independent. However, graphs such as those in figure 2, show that there is a difference in the amount of variation in the time index between different skill levels: more skilled sight-readers show a more constant value throughout a performance than novices.

Another method of measuring the eye-hand span is to calculate the number of notes between execution position and fixation position (the 'note index'). This separation has been more frequently utilised as the measure for the eye-hand span (e.g. Sloboda, 1974). All subjects look ahead in the music, compared to performance, just as readers look ahead when reading aloud. For both tasks, this separation can be grossly demonstrated at the end of a page, where the performer/reader will turn the page before performance has reached the last note/word. For this study, 'note units' were used to calculate the note index, where one note unit was equal to all the notes performed simultaneously, by both hands, irrespective of note length.

In contrast to the time index results, the note index is more obviously skill dependent. Skilled sight-readers were found to show mean note indexes of about 5 note units, whereas novices looked only 2 note units ahead (see figure 3). However, all subjects showed about the same amount of variation in note index during the performance of a piece.

These results tentatively suggest that musicians can only store information for about a second, independent of skill. This is comparable to the results found for 'eye-performance' measures during other tasks, such as reading (Geyer, 1969), typing (Hershman and Hillix, 1965) and driving (Land, 1996). However, the note index results show that skilled sight-readers can increase the number of notes passing through the buffer in this time.

It has been shown here that eye movements are a powerful tool for the study of music reading. Previous studies have shown that introspection is frequently extremely inaccurate with regards to eye movements and fixation position. In addition, the eye-hand span, in both time and note versions, is very useful for the study of the internal processes which relate information uptake to performance.
Figure 1  Record showing the positions of eye fixations for a skilled sight-reader. Circles indicate fixation position, and adjoining lines indicate saccades. Filled circles indicate that the next saccade moved the fixation point down to the keys of the piano, and are shown linked to the next fixation on the score. New line saccades are shown by the dashed lines.

A. skilled sight-reader

B. novice sight-reader

Figure 2  Graphs showing the time index of two different subjects. Data shown is for 25 fixations of each performance. Dotted lines indicate the mean time index for the entire performance.

A. skilled sight-reader

B. novice sight-reader

Figure 3  Graphs showing the note index of two different subjects. Data shown is for the performance of 50 note units. Dotted lines indicate the mean note index for the entire performance.
References


Musical Psychoacoustics
Roughness of multi-part piano chords
Andrzej Rakowski and Jolanta Taszycka
Fryderyk Chopin Academy of Music, Warszawa

1. Introduction
The sensation of roughness is a major factor contributing to the concept of musical dissonance. Roughness of musical chords is produced by amplitude modulation resulting from beats between partial components of sounds composing the chords structure (Helmholtz, 1863). So far there have been little research showing how roughness or „sensory dissonance” of a chord depends on its musical structure, timbre of its components, tone register, loudness etc. Investigations presented in the present paper may shed some light on the problem of sensory dissonance in piano chords.

2. Experimental methods
Relative roughness was assessed in chords containing 4, 5, 6, 7, 8, 9, 10, 11 and 12 tones played mezzo forte on an upright piano „Schimmel” carefully tuned. Students of the Sound Engineering Department of Chopin Academy of Music served as subjects. All of them had passed at least 10 years’ musical training including a 1- or 2-years’ course of timbre solfege („technical listening”). They participated in half-hour listening sessions every week in groups of 4-6. The average number of subjects taking part in each test was 14 and was never smaller than 12. Two psychoacoustical methods were used. In a „Two Anchors’ Method” (Rakowski and Heering, 1979; Taszycka, in press) subjects had to assess roughness of single piano chords by attaching to each of them one of seven numbers within the range from 1 to 7. The numbers 1 and 7 were arbitrarily attached to lower and upper anchors. The anchors were chords whose roughness had been found in a preliminary experiment as the smallest and the greatest within the whole set under investigation. The Two Anchors’ Method was used throughout most parts of the present investigations. Only in two cases another method, a „Method of Free Scaling” was applied. In this methods subjects were allowed to use any positive number while assessing roughness, and no anchors were used.

Choosing actual anchors was an important part of the Two Anchors’ Method. This was done in a preliminary experiment performed separately within each set of chords containing a given number of components (e.g. 4, 5, or 6 etc.). Subjects listened 3 times to a sub-set of chords containing no more than 22 elements and decided which of the chords was most smooth and then, after further 3 listenings, which of the chords was most rough. Elements of the sub-set were randomly chosen from a set which might have been composed of larger number of chords (e.g. the set of six-part chords representing the six-order interval structure contained 79 chords). The partial anchors chosen for each sub-set were then compared in additional experiments and final pair of anchors for a given set of chords (i.e. containing a given number of tones in a chord) was chosen.

In main parts of the experiment the following procedure was adopted. Three chords from a given set were presented in a row: lower anchor, upper anchor, and a chord to be assessed. All chords were played on the piano by the experimenter 1 second long with 1-s silent intervals in-between. There was a 3-second pause after each task intended as a time (very limited) for the subjects to make their decisions and write proper numbers in the answer sheets. Then, immediately the next task started. The subjects were carefully instructed that they should not assess chords for their musical consonance or dissonance. The only auditory sensation they
should respond to was the sensory effect of roughness. The short time in the experimental procedure left for taking decisions was in favor of rising immediate reactions to sensory stimuli.

3. Stimuli
Zalewski’s (1972) Theory of Interval Structures was taken as a base for classifying multi-part chords and choosing the material for present investigation. The interval structure of any set of tones within the equally tempered 12-semitone scale is formed by transposing all components of the set in such a way that they are all within the interval of one octave and allowing the so-formed within-octave set of tones to be further transposed to any place within the musical pitch range. Thus the interval structure or cyclic interval structure as Zalewski calls it, is fully represented by a series of intervals between subsequent tones. The lowest tone of the octave-reduced set is doubled one octave higher, so the number of intervals representing a structure is equal to the number of non-octave-doubled tones of the set. Intervals forming the interval structure are expressed in numbers of semitones and their sum is always 12. The interval structure being fully defined by an ordered series of intervals may be presented in several forms, depending on the interval placed at the beginning of a series; e.g. a dominant seventh chord may be presented in four different forms: (4332), (3324), (3243), or (2433). It has been agreed that the representative form for an interval structure (we call it here a cardinal form) is such, that the smallest interval is written always in the last place, and the largest interval, if possible in the first place. It should be noted that all chords used in the present investigation were constructed in the form mirroring the cardinal form of corresponding structures.

A convenient and very illustrative form of presenting interval structures is their graphic representation in form of a divided circle. In Fig. 1 two interval structures are presented in such a way, a dominant seventh and a minor third chord. Here the cyclic nature of this representation is clearly seen: We may begin scanning the circle clockwise starting from any number.

Fig. 1. Cyclic interval structures of the fourth order: Dominant seventh (4332) and minor third chord (3333).

According to Zalewski’s (1972) analysis there are following numbers of interval structures of the n-th order, (i.e. composed of n intervals or n component tones within an octave):
1 structure for n=1, 6 structures for n=2, 19 structures for n=3, 43 structures for n=4, 66 structures for n=5, 79 structures for n=6, 66 structures for n=7, 43 structures for n=8, 19 structures for n=9, 6 structures for n=10, 1 structure for n=11 and 1 structure for n=12.
As have been said before the present investigation concerned comparative analysis of roughness in chords composed of 4, 5, 6, 7, 8, 9, 10, 11 and 12 within-octave components. The structure of all these chords exactly mirrored the cardinal form of corresponding interval structures. The number of chords in each set was equal to the number of structures of the corresponding order, i.e. there were 43 four-part chords, 66 five-part chords, 79 six-part chords etc. All chords had tone D₄ (293.7 Hz) as their lowest component.

4. Results obtained with the Two Anchors’ Method
As a result of the experiment chords of each set corresponding to interval structures of a given order were organized in sequences of increasing roughness. The rank order of roughness in each chord within a set was computed as a mean value of numbers attached to it by all subjects. The rank order of chords within separate sets is shown in Fig. 2.

![Fig. 2. Ranking of roughness in chords corresponding to n-th order interval structures.](image-url)
As can be seen in the figure the Two Anchors' Method appears not particularly well suited for investigating the range of roughness in chords representing interval structures of various order. The larger the number of tones in the chords the more difficult and ambiguous the ranking of roughness appears. In spite of the fact that the anchors always keep their arbitrarily attained values of 1 and 7, roughness of all other chords becomes less and less differentiated. Coefficients of differentiation $C$ were computed in the form $C = \frac{R_{\text{max}}}{R_{\text{min}}}$, meaning the ratio of most differentiated rank orders of assessed chords excluding the anchors. Their values for chords corresponding to interval structures of 4th, 5th, 6th, 7th, and 8th order was correspondingly 2.86, 2.67, 2.54, 2.08, and 1.96.

The analysis of variance performed on the results of ranking showed that the differentiation of chords with 4 through 8 components was statistically meaningful ($p = 0.01$). For chords corresponding to the interval structures of higher order it was statistically meaningless.

5. Experimental performed with the Free Scaling Method and its results
Additional results were obtained from the supplementary experiment performed with the Method of Free Scaling. In that experiment two sets of chords were used; they were mostly composed of lower and upper anchors previously applied in the Two-Anchors' Method, and supplemented with two-component and three-component chords chosen in a similar way in the course of previous experiments (Rakowski and Heering, 1979). Chords of each set ("lower anchors" and "upper anchors") were separately presented 3 times in random order to a group of 14 listeners. After the third presentation of each chord listeners had to attach to it any positive number that, according to their opinion, would correspond to the strength of its roughness. The results of scaling, averaged and normalized, are presented in Fig. 3.

![Fig. 3](image)

Fig. 3. Relative roughness of within-octave chords representing structures of the order 2 through 10 (having 2-10 components). Lower anchors mean chords with minimum roughness; upper anchors mean chords with maximum roughness. Sequences of digits in brackets represent interval structures; digits indicate sizes of intervals in semitones; symbols 10 and 11 mean ten and eleven.
6. Discussion and conclusions
Ranking of roughness in middle-register piano chords with equal number of within-octave components is easier at smaller number of components. Chords with more than eight components produce roughness which does not depend on the detailed distribution of the constituent intervals.

Interval structures of equal order revealing extremely different roughness (lower and upper anchors) offer some opportunity to analyze factors contributing to the creation of sensory dissonance. The corresponding pairs of anchors (lower anchor/upper anchor) for structures of the order 2 through 8 are the following: (75)/(111), (543)/(1011), (4332)/(9111), (43221)/(81111), (22321)/(711111), (2232111)/(6111111), (22311111)/(51111111). Lower anchors as opposed to upper anchors are as much as possible constructed of the intervals typical of their participation in the harmonic series. In chords corresponding to the structures of higher order such a possibility is very limited; then, in those chords minor thirds (3) and major seconds (2) remain as the only intervals contributing to the decrease of roughness.

The upper octaves, or, in other words, chords selected as most dissonant, have among structures of various order exactly the same shape. They simply cumulate as many minor seconds (1) as possible and fill the remaining octave space with an appropriate single interval. However, it should be noted that according to general assumption all chords used in the present investigation mirrored only cardinal form of any interval structure. The possible difference of experimentally assessed sensory dissonance e.g. in piano chords (9111) and (1119) remains to be found in future investigations.

Conclusion that may be drawn form scaling roughness of lower and upper anchors (Fig. 3) confirm the general finding from the ranking experiment (Fig. 2). The range of scaling is larger within a set of less dissonant chords (lower anchors) than within a set of chords highly dissonant (upper anchors). The other conclusion could be easily predicted: The more within-octave components in a piano chord the more dissonant it sounds.

Acknowledgment
The work was supported by State Committee for Scientific Research, Grant No 1 P107 022 07.

References:
Perceptual constancy of musical instrument timbres; generalizing timbre knowledge across registers

Gregory J. Sandell and Michael Chronopoulos

Parmly Hearing Institute
Loyola University Chicago, USA

Introduction

The notion of *timbre* as encompassing the multitude of possible sounds associated with a source (for example, "the timbre of the violin") is a complex problem for auditory perception and psychomusicology. Listeners somehow learn to identify musical instruments with a single label (violin, oboe, etc.), in spite of the countless varieties of acoustic signals that an instrument produces over different pitches, loudnesses, and expressive inflections. This phenomenon, the perception of an invariant characteristic that remains constant over many transformations is sometimes termed *perceptual constancy* (Shankweiler et al., 1977; Handel, 1989). The perceived formal unity of timbres associated with a source might also be termed the source's *macro-timbre* (Sandell and Chronopoulos, 1996).

One theory of how perceptual constancy for a source is acquired by the listener is through the process of discovering invariants across the source's range of sounds and adjusting for differences between such sounds to form an abstract and stable representation of the source. In the field of speech this process is known as *normalization* (Pisoni, 1997). This conversion entails selectively filtering out certain features of sounds as irrelevant or idiosyncratic so that widely varying acoustic signals can be regarded as equivalent. The source's signals, as filtered by the listener, are then standardized into a kind of prototype or template of the sound. Previous research has shown increased mental processing (slower reaction times) when listeners are required to filter out irrelevant information, which is thought to be evidence of normalization (Mullennix et al., 1989; Cho et al., 1993). An experiment by Pitt (1995) showed that previously ambiguous sounds (mixtures of trumpet and piano) were more clearly categorized as either one or the other following increased exposure to prototypical instances of one of them (e.g. normal trumpet), an outcome consistent with the normalization hypothesis.

Presumably pitch is one of the dimensions of variation over which listeners must learn to reconcile a wide variety of acoustic signals with a single source. In this context, an instrument's macro-timbre encompasses several different levels of sound quality: (1) the sense of a quality that stays constant over pitch, (2) the broader difference in quality from one register to another, and (3) the continuous changes in quality over consecutive chromatic pitches. Learning an instrument (that is, learning to identify it reliably) thus depends on hearing a variety of pitches...
from the source and comparing their timbres. Once learned, a listener can even identify
the instrument from novel pitches (i.e. not previously heard on that instrument).

If this is true, it will be highly advantageous to the learning process to hear sounds in close
temporal proximity, in order to compare their timbres. If sounds are heard at a temporal distance
from one another, the listener will have to rely more on memory to compare them, and thus will
be disadvantaged. An alternative hypothesis is that listeners may need only learn individual
instances, and that little is gained by hearing them in close temporal proximity. These two
possibilities were tested in the study described here.

Stimuli and Procedure

The stimuli consisted of natural recorded tones of the oboe and English horn from the McGill
University Master Samples, using pitches playable by both instruments (c4 to g5); previous
research has shown that the two instruments are confused in this range (Sandell and Sullivan,
1997). Each note was normalized to a standard duration of 500 ms by imposing a 20 ms linear
fade 480 ms into the tone. The pitches were divided up into two sets of ten tones each (see figure
1), producing low register (set 1, c4 to a4) and high register (set 2, a#4 to g5) sets of pitches;
listeners were trained on pitches from one or the other of the two sets. Set 1 listeners were trained
only on the notes c4, d4, e4, f#4 and g#4 (see figure 1a, solid notes); these will be called the
listener's familiar notes; all others are called novel notes. For these listeners the notes close
in pitch to the familiar notes (c#4, d#4, f4, g4 and a4) are called neighbors (figure 1a, unbeamed
hollow notes). The complete set of notes comprising the familiar notes and neighbor notes are in-
register notes. The remaining notes a#4 to g5 comprised the out-of-register notes (figure 1a,
beamed hollow notes). To see if ability to abstract information from one register to another is
symmetric across registers, a second group (Set 2 listeners) were trained with the high pitches as
their "in-register" pitches (see figure 1b).

Within each group (set 1 vs. set 2 listeners), listeners were further subdivided into two more
groups, designated as either single- or multiple-note-trained-listeners. A training trial consisted
of an oboe stimulus followed by an English horn stimulus, or vice-versa; listeners had to report
the order in which the two instruments appeared. For single-note-trained listeners the stimulus in
each interval was one of the five notes in the training set; for multiple-note-trained listeners, the
stimulus in each interval consisted of all five notes played in succession (isochronously, randomly ordered). Single-note-trained listeners learned all five notes over the course of five
different trials; multiple-note-trained listeners heard all five notes in every trial. Feedback was
presented, and their score was continually updated. Listeners were allowed to continue in the
study once they demonstrated near-perfect scores for an extended duration of time. Most
listeners were not musicians and had little prior experience either hearing or distinguishing oboe
and English horn.

Listeners were tested on instrument identification for all pitches (c4 to g5). For all listeners,
a trial consisted of a single note by either oboe or English horn: listeners had to identify which of
the two instruments they had heard. Sixteen listeners participated, four in each of the four groups
(two training types, two training registers). Each listener were presented each stimulus 16 times,
and their responses were averaged.
Results

For each note that was tested, the two scores for oboe and English horn were averaged together into a single score; thus an individual score for a given note reflects the distinguishability of oboe and English horn for that note. For the analyses in this section we introduce the following labels for categories of variables: learning-type (listeners that were trained with single-notes versus multiple-notes); training-register (listeners that were trained on the low versus high register pitch set), stimulus register (responses for low- versus high-register pitches, disregarding training), register membership (responses for in-register vs. out-of-register pitches for the given listener), familiarity (familiar notes versus all novel notes, that is, neighbors and out-of-register notes) and within-register familiarity (familiar notes versus neighbors within the listener’s training register).

The means for the various categories followed an order consistent with the familiarity and registral proximity of the notes in the category. The highest mean went to familiar notes (.827), next to in-register notes (.792), then neighbors (.756), then novel notes (.726) and finally out-of-register notes (.710). The differences between three of these categories can be seen by comparing the smoothed versions of the data in figure 2. There was 17% greater accuracy for familiar notes over out-of-register novel notes \(t(1510) = 8.124, p < .001\), 10% greater accuracy for familiar notes over novel notes \(t(2014) = 7.385, p < .001\), 8% greater accuracy for in-register notes over out-of-register notes \(t(2014) = 6.849, p < .001\), and 7% greater accuracy for familiar notes over neighbors \(t(1006) = 4.530, p < .001\).

The overall effect of training register is shown in figure 3. Although high pitches tended to be better recognized overall, listeners were better at their register relative to listeners trained on the other register. That is, low-register-trained listeners outperformed high-register-trained listeners on low pitches (by 7%), and high-register-trained listeners outperformed low-register-trained listeners on high pitches (by 9%).

There was no superiority for either learning type overall, but learning-type interacted with various other conditions. Effects for learning-type differed from within-register comparisons to across-register comparisons. Figure 4 shows the within-register comparison, comparing scores for familiar and neighbor notes. The superiority of familiar pitches over neighbor pitches is around 7% both for single-note-trained and multiple-note-trained listeners, with no interaction. Thus the advantage for the notes one is trained on over notes close in pitch to those notes is independent of type of training. Figure 5 shows the across-register comparison, comparing scores for in- and out-of-register notes. A strong interaction is found. Listeners with single-note training performed 15% more accurately on their 10 in-register notes than their 10 out-of-register notes, whereas listeners with multiple-note training has consistent performance across scores for the two registers.

Summary and Conclusions

Controlling the proximity of presentation of pitches in training exerted an apparent influence on the relative identifiability of familiar pitches versus novel pitches that were registralement remote from the familiar ones. Hearing notes in isolation has the benefit of maximizing identifiability for
those particular notes, but at the expense of developing a generalized knowledge of the instrument. A listener trained in this way performed considerably worse on new, registrally remote notes, whereas a listener who heard notes in close succession (thus facilitating comparisons between notes) performed equally well on both categories of notes. This suggests that making comparisons during learning allows listeners to abstract information from one register to another.

The two listeners did not differ in the degree of superiority of identifiability of familiar notes over identifiability of notes that were close in pitch. We speculate that this task is more a matter of detecting the similarity between a novel note and nearest note in memory (i.e. from training), rather than making an abstraction.

Further work needs to be done to consider the contribution of other factors of timbre perception in identifying novel notes. For example, listeners could be using distinctive features to abstract knowledge across registers. Many listeners reported that the English horn overall had a breathy and forced or constricted quality; these features were more prominent in high notes, which could explain the overall greater accuracy for high register notes. Also, high scores were found for certain notes, indicating that certain notes were prototypical for the instrument(s). Thus, judgments may have been made by making a direct mental comparison between each incoming note and one that was heard as particularly “characteristic” during training.

Another question to consider in the future is the degree to which a listener’s knowledge of an instrument is affected by the number of training pitches. Previous research shows that identifiability of many instruments is not constant over their entire performing ranges of pitches (see Sandell and Sullivan, 1997); for example the confusability of violin and viola is not constant over the two instruments’ ranges of pitches, even in the region where they overlap. Therefore the importance of the number of pitches could be obscured by the presence or absence of such “characteristic” pitches.

For this reason there might be some advantage to exploring the question of perceptual constancy with more controlled (artificial) stimuli in which individual differences in identifiability from one pitch to the next were limited. However, it remains an open question as to exactly what acoustical features are the source of perceptual constancy; thus it is unclear how one ensures that such sounds would have sufficient richness of timbral change over pitch (or loudness, manner of playing, etc.).

References

Figure 1. Categories of pitches for (a) a “Set 1 listener” and (b) a “Set 2 listener”. Pseudorhythmic values indicate category: familiar pitches (solid-head notes), neighbor pitches (unbeamed hollow notes), and out-of-register pitches (beamed hollow notes).

Figure 2. Identification scores for pitches collapsing across Set 1 and Set 2 listeners, and single- and multiple-note-trained listeners. The lines are smoothed solutions for the corresponding individual datapoints.
Figure 3. Performance on high vs. low register pitches for listeners trained on high vs. low register pitches. ANOVA on training register: $F(1,143) = .303, p = .583$ (n.s.); stimulus register: $F(1,143) = 50.79, p < .001$; interaction: $F(1,143) = 36.418, p < 0.001$.  

Figure 4. Performance on two categories of in-register pitches for the two types of learning. ANOVA on learning type: $F(1,35) = 23.47, p < 0.001$; within-register familiarity: $F(1,35) = 19.90, p < 0.001$; interaction: $F(1,35) = .466, p = .500$ (n.s.).

Figure 5. Performance on the two registers of pitches for listeners for the two types of learning. ANOVA on learning type: $F(1,143) = 1.087, p = .299$ (n.s.); register membership: $F(1,143) = 12.38, p = 0.001$; interaction: $F(1,143) = 48.861, p < .001$.  

Table: Learning Type

<table>
<thead>
<tr>
<th>Learning Type</th>
<th>Single Note</th>
<th>Multiple Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-register</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Out-of-register</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The accenting potential of pitch change has been acknowledged in theoretical treatments (Jones, 1976; Lerdahl & Jackendoff, 1983) and demonstrated in empirical research (Boltz & Jones, 1986; Drake, Dowling, & Palmer, 1991; Monahan, Kendall, & Carterette, 1987; Thomassen, 1982). Pitch change in a context of no pitch change, greater changes in pitch in the one direction, and reversals of pitch change produce the sense of accent that has been observed in verbal reports of subjects and also in perceptual effects that they produce.

One finding that has a relatively long history in music cognition is the perceptual effects related to accenting by a single dimension of sound (e.g., Woodrow, 1909). Research has shown that physical manipulations of one dimension in sound sequences can affect the perception of other dimensions. I have observed such interactions with intensity and pitch as the accenting dimensions and timing and intensity, respectively, as the monitored dimensions in earlier work (Tekman, 1995, in press).

In recent work (Tekman, 1996) I examined the effects of intensity accents on sensitivity and response criterion for detecting variations in timing of sounds. I found that accenting by higher intensity reduced sensitivity to variations in the duration of intervals preceding accented sounds. This effect was greater if accented sounds preceded shorter rather than longer silent intervals and if the accents were spaced regularly rather than randomly. In addition, bias to report variations in timing was greater especially if longer, rather than shorter, intervals preceded accented sounds.

In the present research project the effects of melodic accents on perceived sound intensity were examined in the same manner. Subjects tried to distinguish sequences of pure tones with equal intensity from sequences in which there were variations in intensity. In some conditions of the experiment larger changes in pitch were incorporated in the sequences, which were chromatic scales except for this variation, in order to create melodic accents. The regularity of the melodic accents and whether the deviant tones had higher or lower intensity were the other independent variables. Sensitivity and response criterion were measured in a signal detection procedure.

**Method**

**Subjects.** A total of 48 subjects participated in the experiment. However, some of the subjects could not reliably make the intensity discrimination they were asked to make. The data from these subjects were not used. The reported averages are based on the data from 24 subjects. In selection of these subjects the criterion that a subject should have a sensitivity of at least 1.0 in at least one condition of the experiment was adopted.

**Stimuli.** Stimuli were pure tone sequences of 12 tones each. The unaccented sequences were ascending and descending chromatic scales. The beginning pitches of the ascending sequences were selected out of a range of C₄ to F₄ and the beginning pitches of the descending sequences were selected out of a range of C₅ to F₅. In the accented sequences four of the twelve tones were preceded by a pitch change of three semitones rather than a
single semitone. In the regular sessions these pitch skips were regularly placed before every third tone in the sequence. In the random sessions the position of the pitch skips were randomly selected. In half the sequences in each session all the tones had equal intensity. In the remaining sequences the intensity of four of the twelve tones, the tones following the larger pitch changes in the accented sequences, differed from the rest of the tones by 3 dB. In four sessions this difference was in the positive direction, that is, four tones had higher intensities. In the other four sessions this difference was in the negative direction, that is, four tones had lower intensities. The subjects were asked to indicate for each sequence whether all the tones had equal intensity of some had higher and some had lower intensity.

The experiment included 8 experimental sessions that were factorial combinations of the three independent variables, that is, presence of melodic accents, direction of intensity differences, and regularity of intensity differences. Each session consisted of 72 trials. The experimental sessions were preceded by two practice sessions of 72 trials each. The order of the experimental sessions varied from subject to subject according to a Latin Square.

Results

The average sensitivity and response criterion for each condition in the experiment is given in Tables 1 and 2, respectively. Higher criteria correspond to greater tendency to report that there were no intensity differences and lower criteria correspond to a tendency to report that there were intensity differences. The zero point corresponds to equality of the probabilities of the two types of responses.

Table 1. Mean sensitivity as a function of regularity, accenting, and direction of deviation.

<table>
<thead>
<tr>
<th></th>
<th>Regular</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unaccented</td>
<td>Accented</td>
</tr>
<tr>
<td>Positive</td>
<td>2.20</td>
<td>1.35</td>
</tr>
<tr>
<td>Negative</td>
<td>1.99</td>
<td>0.81</td>
</tr>
</tbody>
</table>

The main effect of regularity was in the direction of higher sensitivity for regular sequences than irregular sequences ($F_{1, 23} = 4.3, MS_e = 0.45, p = .05$). Subjects could detect intensity variations better if the deviant tones were placed regularly rather than randomly. This was in the opposite direction of the effect that was found for detection of timing variations. The main effect of accenting was reducing sensitivity ($F_{1, 23} = 66.48, MS_e = 0.71, p < .001$). This effect was greater for negative deviations than for positive deviations but this effect did not reach significance ($F_{1, 23} = 2.67, p = .116$). The only other significant effect on sensitivity was higher sensitivity for positive than for negative deviations ($F_{1, 23} = 4.56, MS_e = 0.60, p < .05$)

Table 2. Mean criteria as a function of regularity, accenting, and direction of deviation.

<table>
<thead>
<tr>
<th></th>
<th>Regular</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unaccented</td>
<td>Accented</td>
</tr>
<tr>
<td>Positive</td>
<td>0.09</td>
<td>0.20</td>
</tr>
<tr>
<td>Negative</td>
<td>0.08</td>
<td>0.22</td>
</tr>
</tbody>
</table>

The only main effect that approached significance for the response criterion was the effect of regularity ($F_{1, 23} = 3.73, p = .066$). On average, subjects were less willing to report intensity
deviations when the deviations were placed randomly. However, there was a very reliable interaction of the effects of accenting and regularity ($F_{1,23} = 9.89, MSe = 0.04, p = .01$). This interaction indicated that the difference between regular and random sequences was limited to the unaccented sequences. Accenting raised the response criterion for regular sequences whereas it lowered the criterion for random sequences. Thus, the response criteria for all types of accented sequences were comparable.

**Discussion**

The results of this experiment are radically different from the results of the analogous experiment in which accenting was achieved through manipulations of intensity and detection of deviations in timing was investigated (Tekman, 1996). In that experiment better performance for random deviations than for regular deviations was observed, especially for accented sequences. In the present experiment subjects found detecting deviations in regular sequences easier than detecting deviations in random sequences and this difference was not affected by accenting. There was a tendency towards greater reduction in sensitivity for negative than for positive deviations as a result of accenting, but this effect did not reach significance. Such an interaction would have provided evidence that tones that followed larger changes in pitch were perceived as having higher intensity. It is possible that such an effect, which was observed in other work (Tekman, in press), could not be detected because of large individual variations.

The results about response criteria were not similar to the results of the earlier experiment (Tekman, 1996) either. Consistent with longer perceived intervals preceding higher intensity tones, in the earlier experiment subjects were most willing to report timing variations in accented sequences with deviations in the positive direction. This effect was not replicated in the present experiment. The most reliable effect related to response criteria was an interaction of accenting and regularity. Although subjects were much more willing to report intensity differences with regular rather than random deviations without melodic accents, this effect completely disappeared when the accents were present.

The differences between the result of the present experiment and the earlier experiment may indicate that accenting affects perception of intervals separating perceptual groups in a way that is quite different from how it affects perception of elements that belong to perceptual groups themselves. One important caveat is the very high rate of subjects whose data could not be used in the present experiment. The unexpected difficulty of the task may be due to use of an inappropriate magnitude of deviation. This is not much likely in the light of the fact that in earlier research (Tekman, in press) the magnitudes of deviation used in the two experiments proved to be comparably difficult, albeit in a different subject population. Another possibility is that the subject could not develop a clear idea about which dimension of sound they should concentrate on from the instructions. In fact, during debriefing subjects displayed considerable variation and inconsistency in terms of the terms they used to refer to different dimensions of sound. Possibly the task could not be made sufficiently clear to the subjects because the Turkish language does not provide consistent terminology for such communication. In any case, repeating this experiment with changes that would ensure reliably better performance from the subjects appears necessary.
References


Ethnomusicology
Observations on some correspondences between ethnic music and animal calls.

Inge Cordes, Universität Bremen, FB 9

My research is intended to examine if the melodic contours which young children universally learn during initial speech acquisition are also of importance for music and music comprehension. These melodic contours are described by Papoušek et al. (1989, 1991). These authors have also found that the composition of melodic contours used by childminders varies with social context. For my investigation, four social situations were selected and corresponding songs collected. These are: praise songs, war songs, songs to stimulate attention and lullabies. All songs stem from a variety of cultures and have been transmitted orally (ethnic music). Subsequently their melodic construction was analysed. In fact the result shows significant differences in the composition of melodic contours between the four categories of songs. However, this is not the main concern now, instead, our attention will be focused on the so-called 'biological parameters'. They comprise duration, the number of different shapes of contours used within a given song, and the average duration of the melodic contours of a song which have also been established.

As table I shows, on average lullabies are the shortest of all the songs investigated. This has certainly to do with infants’ limited attention span. Praise songs are longest. People can apparently endure praise for some considerable time. The second column, too shows convincing connections. So, it is easy to understand why lullabies, which are thought to be mainly soothing, are composed of the smallest number of different melodic contours. The other song categories contain a larger number of different shapes of melodic contours, being less monotonous, with the exciting war songs taking the lead. The data of the last column reveal a clear difference in the average duration of the melodic contours; the melodic contours of the war songs are rather short as compared to those of the other song categories which, by contrast, all belong to pleasant situations.

This finding is of importance. Tembrock (1971), a German ethologist, states that in the acoustic system of communication the control of distance is most important. He claims that distance-reducing calls have an affinous effect. By contrast, calls which widen the distance between individuals have a diffugous effect. Affinous calls are characterized by a relatively long rise time until they reach full amplitude and by a longer temporal extension. This is achieved by stretching the sounds, by frequent repetition of calls or by a series of calls. In vertebrates, they have a tonal character with dominating frequencies. Diffugous calls reach maximum amplitude quickly and are short. They are not or only irregularly repeated and are of a noisy character. The broad spectrum of frequencies prevents or restricts an adaptation in the percipient and causes the individual to seek refuge, often in flight.

<table>
<thead>
<tr>
<th>Song-category</th>
<th>Average duration of song /sec</th>
<th>Average number of diff. shapes of contours</th>
<th>Average duration of melodic contour /sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Songs to arouse attention, n=22</td>
<td>108.32 (71.28)</td>
<td>3.73 (1.32)</td>
<td>5.14 (3.28)</td>
</tr>
<tr>
<td>Lullabies, n=63</td>
<td>85.34 (46.6)</td>
<td>2.57 (1.11)</td>
<td>4.51 (2.29)</td>
</tr>
<tr>
<td>Praise songs, n=42</td>
<td>138.67 (70.08)</td>
<td>3.83 (1.32)</td>
<td>3.99 (2.25)</td>
</tr>
<tr>
<td>War songs, n=38</td>
<td>125.08 (46.46)</td>
<td>4.11 (1.06)</td>
<td>2.50 (1.64)</td>
</tr>
</tbody>
</table>

Table I Biological parameters of the four song-categories. Standard Deviation is given in brackets.
As an example Tembrock cites Spittka who has been able to show that rats, having the opportunity to choose between different sounds, prefer those with long rise times, of longer duration and with rhythmic repetition. The findings of Papoušek et al. (1990) emphasize that these principles are meaningful for humans, too. They have shown that while expanded melodic contours manage to attract the attention of four-month old infants, a short rising-falling contour fails to do so. In this context, the results of Gabrielsson’s investigation (1995) concerning a folksong played in different emotional versions one must also mention. The angry version is played fastest and the onset of the first tone is extremely rapid as compared to the other versions. This indicates that the above principles are also effective in our music.

Let us now take a look at particular songs. The Ponca War Dance (Canadian Indians) is an example of a song belonging to aversive behaviour, fig. 1. The analytic graph shows it to be composed of prevailing short, steeply declining melodic contours and only some sinusoidal figures which also decline rapidly. The average duration of the contours in this song is 1.77 sec, which is very short indeed. The song is accompanied with bells and drums, which underline the beatlike sounds sung in chestvoice. A second example of a song composed of diffusional melodic contours is a Paddle Dance from Tahiti, fig. 2. It also contains many short and steeply declining figures. They are sung in a noisy manner, probably accompanying the beating of the paddles intended to drive fish into a net. The average duration of the contours in this song is only 1.73 sec.

Of course, there are other forms of utterings in aversive contexts, too. This shall be demonstrated by the Iroquois War Dance (Canadian Indians), fig. 3. It starts with rapidly rising contours which are long and interrupted by large leaps, all of which giving the impression of power. This song is also sung in chestvoice, which, following Jürgens (1986), is used for intimidation. Trojan (1982) links it to impressing behaviour, often occurring in anger to frighten the adversary. This kind of aggressive utterance with longer contours predominates in this song; therefore the average duration of its contours is above average for war songs. Given that Scherer (1982) claims the phylogenetical continuity of emotions, it may be allowed to consider the behaviour of closely related animals. For chimpanzees it is reported (Vogel 1994) that corresponding sounds in charging display, following Goodall, can be longer or shorter, with ape making himself as large as possible and the hairs of his coat even bristled up. For Rhesus monkeys (Ploog 1982) has been shown a whole continuum of diffusional utterings with different amounts of aggressiveness. A long roaring sound which is probably comparable, is uttered only by very self-confident animals. Growling is another diffusional sound, uttered by animals which are only slightly alarmed. It is composed of mainly deep frequencies. A corresponding example from a song will be shown later. In this context it is of interest that La Barre (1954) claims that Gibbons, whose acoustic communication is most developed, achieve the same emotional state in the entire group by uttering calls corresponding to a given situation. This is of particular advantage, when they are about to attack another group. Sometimes it may even render fighting unnecessary, because the others have withdrawn because of the racket. The development of human war songs and dances may have taken place in an analogous manner until they became rituals.
Tembrock (1982) states that acoustic communication takes place in different fields: 1. in the near field, where permanent communication can occur, 2. in the transition field; and 3. in the distance field. Difficult calls will occur in the near field with direct contact like growling, or in the transition field to demarcate or to defend territory. Affinal calls occur either in the near field or in the distance field to call for someone very distant or as alarm calls. Tembrock characterizes these long-distance calls by intense temporal stretching and high sound intensity. As a protection against disturbances and mistakes in the transfer they are rich of redundancies. One example is an Ancient Sentinel Alarm from New Zealand, fig. 4. It is uttered by a guard when on sentinel duty sitting in a box of reed atop a tall column aside the main entrance into the fortified village. A second example is Addham Muezzin’s Call to Prayer from the Middle East, fig. 5. Both analytic graphs show very long drawn out melodic contours often staying on one level. Sometimes they are ornamented, possibly as a way of prolonging a figure, while at the same time interrupting its monotony. The average duration of the contours, 5.03 sec and 7.39 sec resp., is very long. Most of the contours of both songs start rising, with those of the alarm song more rapidly.

In the near field with permanent communication many different sounds are uttered. Those belonging to pleasant contexts are playsounds and sounds of cosiness for example and above all sounds of contact to transfer latent or current information. They are addressed either to the whole group or to special partners who are originally mother and child. Lullabies as well as praise songs are most similar to these sounds. One example is the first part of Kona Kai opna i ka la‘i, fig. 6. This chant tells of the beauty of the famous district of Kona on the Island of Hawaii. Its softly rising bell-shaped figures, prevailing in this song, are stretched, the contours having a duration of 3.44 sec on average, and it is obvious that there is much repetition. A section of the Song in Praise of Captain George O’ Malley from West Ireland is another example, fig. 7. In this song a succession of different melodic contours of only 2.69 sec on average is repeated twelve times with only little modification. A third example is a Song to Glorify the Ancestors of Borneo, fig. 8. It starts with a very long contour, sung very loud, similar to long calls of the distance field. Then there are longer melodic contours and also shorter ones which follow each other very rapidly, forming a series. This song, besides sinusoidal melodic contours, also contains many bell-shaped contours which are typical of praise songs. So, in the song both elements, praise and calling for attention, are combined. This stands to reason, because the ancestors to be praised are a long distance away. An average duration of the contours of 5.9 sec fits this purpose well.
The lullabies also have temporally stretched melodic contours on average as shown in table I. But things are not as simple as they may appear at first glance. An analysis of the songs has revealed differences concerning the composition of melodic contours. A test then proved that there are at least four types of lullabies, which also differ in their 'biological parameters'.

<table>
<thead>
<tr>
<th>Song-type</th>
<th>Average duration of song/ sec</th>
<th>Average number of diff. shapes of contour</th>
<th>Average duration of melodic contour/ sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swingtype, n=22</td>
<td>90.09 (45.94)</td>
<td>2.05 (0.70)</td>
<td>5.37 (2.73)</td>
</tr>
<tr>
<td>Praise song-type, n=57</td>
<td>89.62 (43.18)</td>
<td>2.68 (1.03)</td>
<td>4.47 (2.30)</td>
</tr>
<tr>
<td>Playsongs, n=17</td>
<td>56.00 (28.69)</td>
<td>2.41 (1.18)</td>
<td>3.87 (1.09)</td>
</tr>
<tr>
<td>Songs of rejection, n=5</td>
<td>116.40 (81.32)</td>
<td>4.40 (0.89)</td>
<td>2.28 (1.04)</td>
</tr>
</tbody>
</table>

Table II Biological parameters of the four types of lullabies. Standard Deviation is given in brackets.

There are very soothing lullabies, called 'swingtype', fig. 9 (Brazilian Indians), with long contours and only little change in the shapes making for monotony. They are probably sung to very young infants. Another group, named 'praise song-type', fig. 10 (West Ireland), shows much similarity with the praise songs for grown-ups, though the average duration of contours is even longer. Besides, the contours are all nicely separated, which is typical of lullabies in general. This may make it easier for the infant to take them in. Furthermore there are playful, entertaining songs, called 'playsongs', fig. 11 (Amazonas, Indians). They are the shortest of all. This may correspond with infants’ limited ability to pay attention which these songs require in particular. Surprisingly there were also found songs with diffusus melodic contours. With them unwanted behavior shall obviously be refused. This is demonstrated by a section from a lullaby sung by a young girl to her babysister. When the baby starts crying, there suddenly emerge steeply declining falling figures, similar to those in war songs, alternating with softly falling contours, fig. 12 (Africa, Yaka). Finally a song will be shown which is of a threatening character. The singer imitates several times the crying of a baby, but before and in between there are long, flat sinusoidal contours, fig. 13 (American Indians). They are sung in chestvoice and seem to correspond to the slightly aggressive warning character of growling which is known from several mammals. These 'songs of rejection' make up only 5 of a total of 83 analysed lullabies. This small number agrees with the findings of Bozzo et al. (1944) and Papoušek et al. (1991) in linguistic research. Songs of pleasant contexts, mainly the 'praise song-type', predominate. Nevertheless, in the field of lullabies the same principles of acoustic communication are at work.

The examples show that Tembrock's general characterisations of animal acoustic behaviour apply also to many features of their melodic design. These correspondences suggest that at least some features of our music making and music comprehension are very deeply rooted in our phylogenetic heritage.
Discography

Anthology of Brazilian Indian Music. Ethnic Folkways Library FE 4311. Javahe, Lullaby-Song, mother and child (Nr. 8, S. B). Figure 9

Authentic Music of the American Indians. Everest Records 3450/3. Ponca War Dance (Platte 1, S. B, Nr. 1) Figure 1

Folk Music of the United States, Songs from the Iroquois Longhouse. Archive of Folk Song, The Library of Congress Music Division. The Iroquois War Dance - wa saa’ se’ (S. B, Nr. 2). Figure 3

Hawaiian Chants, Hula and Love-Dance Songs. Ethnic Folkways FE 4271. Pa mai ka makanil - a Paddle Dance from Tahiti (S. B, Nr. 4). Figure 2

Hawaiian Chant, Hula and Music. Folkways Records FW 8750. Kona kai opna i ka i’a (S. B, Nr. 11). Figure 6

Indian Music of the Pacific Northwest Coast. Folkways Records Album No. FE 4523 (1967), USA. Cradle Song, sung by Mungo Martin (Nr. 4, S. D). Figure 13

Maori Songs of New Zealand. Ethnic Folkways Library P 433. He Wha ka a raarara - Ancient Sentinel Alarm (S. B, Nr. 4). Figure 4

Musique de la haute forêt amazonienne. Tumac-Humac BAM LD 314. Éditions de la boîte à musique, Paris. Berceuse (S. B). Figure 11

Musique Dayak Borneo. Collection Musée de l’homme LDM 30108. Bakung - Song to Glorify the Ancestors (S. A, Nr. 4). Figure 8

Religions of the Middle East. Argo Record & Co, ZRG 3220, ZFB 54. Addhan-Muezzin’s Call to Prayer (S. B, Nr. 1). Figure 5

Songs of Aran, Gaelic Singing from the West of Ireland. Folkways Ethnic Library, FM 4002/ Monographic Series. 1) Seo-Thin seo, s tú mo leath. Figure 10
2) A Song in Praise of Captain George O’Malley (S. A, Nr. 8). Figure 7

Yaka Musik Nr. 8, Africa. Tervuren 8, Belgium. Su, su, e, su, su (Nr. 6). Figure 12
Selection of references


Inge Cordes
Butlandsweg 10
D 28357 Bremen
Germany
1. Introduction

This paper constitutes an exploration of the relationship between word-level prosody and musical rhythm in four Estonian folksongs from the northeastern dialect area, produced by the folksinger Liina Kaskmann and recorded in 1937 (Laugaste 1989). The recordings were analyzed acoustically at the Department of Linguistics of the Ohio State University.

The primary aim of the current study was to study the influence of position within the line on the realization of word-level prosodic patterns. The Estonian folksong line is traditionally labelled a trochaic tetrameter. The line consists of four disyllabic metric feet; alternatively the line can be described as consisting of eight positions, of which four are odd-numbered (1, 3, 5, 7) and four are even-numbered (2, 4, 6, 8). The odd-numbered positions carry metrical and musical stress. In the classical Estonian folksong, first syllables of words in the short quantity are excluded from odd-numbered positions, even though they carry word-level stress. In the strict realization of the pattern in the folksong, the first syllable of such words must occur in an even-numbered position, and a conflict (or competition) arises between word-level stress and musical rhythm. The question we explored concerns the realization of lines in which such a conflict takes place.

In several earlier studies (Ross and Lehiste 1994a, 1994b, 1995) we treated the realization of the Estonian word-level prosodic structure in Setu funerallaments. In the Setu dialect as well as in most other Estonian dialects and in the literary standard, word-level prosody involves a three-way system of durational oppositions, manifested not only in the duration of the stressed syllable, but also in the duration ratio between the stressed syllable and the syllable following it in the disyllabic sequence. The results of our analysis revealed extensive neutralization of word-level durational oppositions.

The songs analyzed during this phase of our study come from an Estonian dialect area that is described as having only a two-way durational opposition at the word level (short and long, rather than short, long, and overlong, as is the norm in most Estonian dialects including Setu). The second aim of this study was to test the manifestation of the prosodic structure of words as produced by the singer, to verify that there is indeed no differential treatment of words that are analyzed as belonging to either the long or overlong category in normative cases, i.e. in most Estonian dialects.

2. Material and methods

The four songs analyzed in the course of the study consisted of a total of 152 lines. This total includes 93 regular trochaic lines, 25 lines we refer to as dactylic, 3 lines that contained more than eight syllable-notes, and 31 lines that contained punctuated notes, 27 such lines being basically trochaic, and 4 dactylic. Dactylic lines contained at least one trisyllabic foot, which resulted in disrupting the agreement between word stress and musical
stress. The material was analyzed using the Kay Elemetrics Computerized Speech Laboratory (CSL); measurements were made of the duration of all syllable-notes as well as the pauses between individual lines. The overall average durations of metric feet in 93 regular trochaic lines, expressed in milliseconds, are given in Table 1.

Table 1. Average duration of syllable-notes in lines consisting of four trochaic metric feet.

<table>
<thead>
<tr>
<th></th>
<th>1-2</th>
<th>3-4</th>
<th>5-6</th>
<th>7-8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>310</td>
<td>279</td>
<td>319</td>
<td>285</td>
</tr>
<tr>
<td></td>
<td>318</td>
<td>278</td>
<td>326</td>
<td>264</td>
</tr>
</tbody>
</table>

The average duration of syllable-notes in odd-numbered positions within the line (1, 3, 5, 7) was 318 msec, that of even-numbered syllables -- 276 msec. The results indicate that even though the traditional notation uses notes of the same duration for both odd- and even-numbered positions, actually the first (stressed) member of the odd-even pair is noticeably longer than the unstressed member.

3. Manifestation of word-level prosodic structure
3.1. Disyllabic words

The contrast between word-level prosodic patterns (short, long, and overlong quantity) is manifested both by the duration of the syllables and by the ratio between the stressed syllable and the unstressed syllable following it within the same disyllabic unit (metric foot). In singing, it is to be expected that the duration of the syllable-notes is determined largely by musical rhythm, which may or may not influence the realization of the word-level patterns. A tabulation of average syllable-note durations for words in short contrastive quantity (Q1) is given in Table 2, which also contains average syllable-note durations of words that would be assigned to Q2 (long) and Q3 (overlong) quantities in the literary standard.

Table 2. Duration of syllables occurring in disyllabic words starting in odd-numbered positions (word stress coinciding with metrical stress).

<table>
<thead>
<tr>
<th>Word type</th>
<th>N</th>
<th>Odd-numbered syllable</th>
<th>Even-numbered syllable</th>
<th>S1/S2 ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>64</td>
<td>293</td>
<td>297</td>
<td>0.987</td>
</tr>
<tr>
<td>Q2</td>
<td>100</td>
<td>320</td>
<td>280</td>
<td>1.143</td>
</tr>
<tr>
<td>Q3</td>
<td>182</td>
<td>328</td>
<td>269</td>
<td>1.219</td>
</tr>
</tbody>
</table>

The table reveals a characteristic feature of the Estonian word-level quantity system: there is a compensatory relationship between the durations of the first and second syllable, so that increasing duration of the first syllable is accompanied by reduction in the duration of the second syllable. Nevertheless, the ratios do not quite reach the extent of those found in the spoken form of the language, where the S1/S2 ratio for Q1 is approximately 0.67, the ratio
for Q2 is 1.5, and that for Q3 is usually somewhat higher than 2.0. It should be remembered that the northeastern dialect is traditionally described as having no opposition between long and overlong quantities, and that the words were assigned to Q2 and Q3 according to what is expected in the literary standard. The results of our experiment show a tendency to converge towards the standard pattern.

3.2. Dactylic lines

We used the term 'dactylic' to refer to lines in which word stress and musical stress diverge, since the cause of the disagreement was primarily the presence of trisyllabic (dactylic) words. The average duration of syllable-notes in the eight positions of dactylic lines is given in Table 3.

Table 3. Average durations of syllable-notes in dactylic lines.

<table>
<thead>
<tr>
<th></th>
<th>1-2</th>
<th>3-4</th>
<th>5-6</th>
<th>7-8</th>
</tr>
</thead>
<tbody>
<tr>
<td>301</td>
<td>298</td>
<td>326</td>
<td>277</td>
<td>329</td>
</tr>
<tr>
<td>289</td>
<td>330</td>
<td>311</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The average duration of syllable-notes in odd-numbered positions within dactylic lines was 322 msec, that of even-numbered syllables – 294 msec. Thus the difference in duration between syllable-notes in odd- and even-numbered positions was smaller in dactylic lines (28 msec) than in trochaic lines (42 msec).

There were 13 disyllabic Q1-words in dactylic lines occurring in even-odd positions; the average durations of the syllables were 266 and 345 msec (the average durations of syllables in disyllabic Q1-words having been 293 and 297 msec, respectively). When a disyllabic word occurs in the even-odd position, the first syllable bears linguistic stress, while the second syllable bears musical stress. The duration of the second syllable here is considerably greater than the overall duration of odd-numbered syllable-notes (i.e. syllables carrying musical stress), which was 318 msec in regular trochaic lines.

There were 29 trisyllabic Q1-words that occurred in final position in dactylic lines, the stressed short syllable falling in position 6 – a musically unstressed position. The average durations of their syllables were 285-323-303 msec. The durational pattern is similar to what was found for disyllabic words occurring in even-odd positions in dactylic lines, but not quite as extreme: the difference in duration between the first and second syllable in trisyllabic Q1-words was 38 msec, whereas in disyllabic words it was 79 msec. (In the whole material, the difference in duration between odd-numbered and even-numbered syllable-notes was 42 msec.)

The whole material contained only two trisyllabic Q1-words occurring in odd-numbered positions (i.e. in a situation excluded by the folksong metrical rule that prohibits stressed Q1-words in odd-numbered positions). The average duration of syllable-notes in these words was 270-353-323 msec, which is comparable to the durations found in trisyllabic words starting in position 6. These words thus seem to follow the word-level durational pattern rather than the pattern predicted from the musical rhythm.
The 14 trisyllabic words with a long initial syllable, or non-Q1-words (starting in odd-numbered positions, where word stress coincides with musical accent) had durations of 331-289-314 msec; the pattern of the first two syllables is comparable to that of disyllabic feet in odd-even positions.

Only one non-Q1 trisyllabic word occurred in line-final position; its syllable-note durations were 315-290-290 msec. As was the case with the two trisyllabic Q1-words starting in odd-numbered positions, the word-level pattern appears to dominate over the expected musical pattern.

4. Punctuated notes

In two of the songs, the singer used lengthening that had been transcribed with punctuation of the stressed note. In one song there were 6 such notes occurring on words in which overlong quantity is expected in the literary standard. In the other song, lengthening of a long syllable occurred in 13 instances, but in 7 instances, it was a Q1-word whose contrastively short syllable was produced with lengthening. The average durations of the two syllables in trochaic feet starting with a punctuated note were 397 and 266 msec for words in Q1, and 403 and 210 msec for words with a long first syllable. In all but three cases, the punctuated note was the highest note in the melodic line.

5. Conclusions

The study revealed a complex interaction between word-level prosody and musical prosody. First of all, musical stress was shown to be reflected in the greater duration of notes occurring in odd-numbered positions. Within this framework, the linguistic word-level quantity patterns could still be detected as deviations from the musical rhythm. There was a clear differentiation of the short (Q1) quantity pattern from the long quantity, and even within the long quantity there was a slight tendency toward differentiation comparable to the distinction between Q2 and Q3 in the literary standard.

The rhythmic pattern imposed by the musical structure influenced Q1-words in particular by shortening the second syllable of disyllabic Q1-words, which is 'half-long' in spoken language. Interestingly, when a trisyllabic Q1-word occurred in line-final position (with its stressed syllable falling into a musically unstressed position), the S1/S2 ratio was manifested in a way characteristic of the spoken language. But the use of punctuated notes seemed to be motivated by melodic requirements, and it had a contradictory effect: adding extra length to the first syllable, the characteristic S1/S2 ratio was enhanced for words that would be classified as overlong in the literary standard, but it distorted the realization of Q1. Still, the realization of trisyllabic words in line-final position provides evidence of the phenomenon called scansion in the description of Estonian folksongs.
6. References


The Image of the Foreign Music in Perception of the Kazakh Kyuishis
Saule Utegalieva
Almaty Kunnangazy State Conservatory, Kazakhstan

Introduction
In repertoires of many Kazakh musicians of 19th and 20th centuries, those pieces have been preserved, which were created in style of foreign music. They have certain titles and special intonations and sounds. Among them single examples are seen as "Kalmak bi" ("Kalmak dance"), "Kamarinskaja by Kazakh" (Russian dance in Kazakh version), which have far similarity with the origins, and groups of pieces are also performed characterised by stable stylistic indications, for example "Turkmen kyulery" ("Turkmen kyuis").

The author studies some perception's mechanisms of foreign music, taken as example of Kazakh kyuishis' creation (the performers-creators of instrumental pieces - kyuis). I examine this phenomenon in the nineteenth and twentieth centuries from an ethno-historical perspective. In study of foreign musical image I mark some aspects: (1) preconditions for its appearance, (2) content and its meaning in the development of the ethnomusical interconnections.

Examination of such a slippery subject as musical perception demands the use not only of ethnomusicological data, but of a variety of disciplines, such as history, ethnography and the terminological tools of psychology. I use a variety of materials, including: published scores of kyuis and own recordings of pieces, collected in field works to Atyrau and Mangystau regions in Western Kazakhstan (1985–87) and Russia's Astrakhan region (1990).

Textual analysis of musical recordings and interpretation of the kyuis reveals the perceptual peculiarities of foreign music. Therefore, one must use analytic methods of research, which allow partial reconstruction and understanding of the creative process's norms. My attempt to separate the national qualities of foreign musical image brings attention to the methods of comparative study, which offer an opportunity to determine "influence or reciprocity", highlighting the historical proximity of these peoples" (V.Zhirnmunskii 1979:68).

1. Preconditions to the advent of the Foreign musical image

I explore a few of the preconditions to this phenomenon's appearance. Socio–historical preconditions. Among the many factors that to varying degrees promoted development of multidimensional contacts, here I will focus on the territorial proximity, the neighborhood as well as together live various peoples.

The author examined two marginal zones of active intercultural interaction in the West Kazakhstan. It is so-called kazakh–tatar–nogai area, located in the northwest of Kazakhstan, on the border with Russia (the territories of Ural, Atyrau and Astrakhan regions) and kazakh–turkmen area in the southwest of the Republic (Mangystau region).

The Lower Volga region is a major multiethnic region. At present time here live not only Russians, but Kazakhs, Nogais, Tatars, Kalmyks and other peoples. Separate colonies of
Kazakhs are at Astrakhan partly Volgograd and Saratov regions off Russia. It was Kazakh territories in Past.

Mangystau was famed as a frontier settlement zone of mainly two peoples: Kazakhs and Turkmens. Alongside Turkmen tribes roamed Kazakh clans, including Adai, Tabyn. Some Adai clans, so-called Turkmen-Adais live on the territory of Turkmenistan.

Cultural preconditions. These pieces arose as a result of old interethnical and cultural contacts with neighboring peoples (Russians, Turkmens, Kalmyks, Tatars and others). Familiarity with the customs and traditions, musics of various peoples were acquired at all kinds of performances, celebrations and in markets, which organized in cities and trade centres such as Urda, Khan Bazar, Khiva, Merv and others (A.Zhubanov 1958: 29,34,47). At these events a variety of competitions occurred, including musical–poetic performance (A.Zhubanov 1958: 222). Here participated not only Kazakh, but Uzbek, Turkmen, Kalmyk musicians. Folk tradition preserved a multiplicity of legends about the competitions of famous Kazakh kyuishis such as Dina, Oskenbai with great Kalmyk musicians and Turkmen sazanda and bakhshi (K.Sydikov 1973: 156). The kyushi's school of Oskenbai (1860–1925) formed in Mangystau, who specialized in the composition of Kazakh paraphrases on the Turkmen themes. In repertoires of these musicians many Kazakh kyuis perform in Turkmen versions. Among them is kyui "Aksak kulan– Zhoshy Khan".

2. Image of the Foreign music in perception of the Kazakh kyuishis. Content and its meaning.

The pieces represent the interest as an attempt of Kazakh musicians to reflect principles of foreign music. We have in view also the degree of accordance of compositional paraphrases with original examples. The musicians, composing the pieces, on the one hand, quoted the legends, histories of the creation, the titles of the pieces, some motifs from the origins; on the other hand, they improvised them and showed their interpretation according to their creative treatment of listening to music. The boundaries between quotation and "transinterpretation" become relative for conditions of oral tradition.

In this work the author shows that many Kazakh musicians are good experts and have extraordinary listening ability. The image of foreign music appearing in their perception has clear characters and corresponds with the original models. Some special peculiarities of the music can be exaggerated (1). For example, in turkmen pieces Kazakh kyuishis use a complex rhythmic stereotypes, the forms with two zonal structures and special modal structure with augmented second interval of Turkmen music, which is called "gyrklar". They fasten the first perne (or mode) on the dombra (a two stringed plucked instrument of Kazakhs), which is called "Turkmen Perne" and is used in Turkmen kyuis.

Comparative analysis shows that Turkmen sazlar are more complex than Kazakh examples.

In foreign music the musicians indicate characters, which are related with their own musical culture. That is, the characters are not only imported but also national ones (2). Kazakh and Turkmen pieces show the similarities the compositional forms, timbre of the dombra and dutar (two stringed chordophone), separate methods of play.

Principles of music development show Kazakh national quality of the pieces. Almost all pieces have a stable compositional structure, characteristic for dombra kyuis of the West Kazakhstan (3). Intational similarities of the pieces with the origins are neutralized by timbre
peculiarities of dombra. Modern Kazakh dombra with long neck and chromatic scale has a diapason of the two octaves. This instrument preserved own timbre in despite of the improvement. The timbre of dombra mainly carries Kazakh nationality (4).

References

Amanov, Bagdaulet

Aravin, Piotr

Gullyev, Shakhym

Karutz, Richard

Nazaikinskii, Evgenii

Sydikov, Kabibulla

Uspenskii, Vladimir and Beliaev, Vladimir.

Zarukhova, Sara

Zataevich, Aleksandr
1931 500 kazakhskikh pesen i kiuev. Alma–Ata: Narkompros KazSSR.

Zemtsovskii, Izali

Zhirmunskii, Victor

Zhubanov, Akhmet
Music Cognition: Melody and Rhythm
This is the first in a series of experiments aimed at studying the process of music perception as directly as possible to discover the perceptual mechanisms involved. The principal rationale behind this work is that we need to gain a deeper insight into the processes that are performed while listening to music. Since music is a serial stimulus that is incrementally processed, these processes should be readily accessible to investigation. The gating paradigm, in which fragments of increasing length are presented and subjects provide responses after each fragment, seems to be the optimal experimental format for this investigation.

In this exploratory study subjects listened to a large number of short tone sequences and indicated of each sequence whether in their opinion it formed a musical percept. The purposes of the experiment were 1) to determine whether subjects can reliably distinguish between tone sequences that lead to musical percepts and those that don't; 2) to test the suitability of the employed experimental paradigm; 3) to arrive at hypotheses concerning the perceptual mechanisms that underlie the formation of musical percepts.

Method

Stimulus material. Stimuli were selected from the 120 permutations of the tone set C4 E4 F#4 G4 Bb4. This set of tones was chosen because we believed that it would be most suited for studying the on-line processes in music perception. Firstly, the tones of the set do not belong to one major key or scale (C E G and Bb are all members of the key of F major, but F# is not), although they all belong to the key of G minor. Secondly, different subsets of tones within the set may evoke the two dominant seventh chords C7 and F#7 (the tones C E G and Bb are elements of the C7 chord, while the tones E A# (Bb) and F# are elements of the F#7 chord) that are the dominants of respectively F and B which are remote keys located at opposite sides of the circle of fifths. Thus the sequence is musically quite ambiguous as a result of which percept formation will greatly depend on the order of the tones in the sequence (i.e. on the permutation used) as well as on the induced key. This is, of course, precisely what we need: a set of stimuli that differ greatly in the extent in which they lead to musical percepts. In a small pilot experiment four musically experienced subjects listened a few times to all initial fragments of sizes 3, 4 and 5 of all 120 series (comprising 5 tones) and judged whether they formed a musical percept. In this way all 360 fragments were provisionally classified as either or not being musical fragments. Based on this classification 53 of the series were selected as stimuli in the experiment. See Table.

Table 1. The 53 stimuli used in the Experiment. All notes are in the fourth octave (USA Standard octave designation). The last three columns (labeled 3-percept, 4-percept, and 5-percept) indicate the number of subjects that judged the initial fragment of that size a musical percept (left sub-column), and the perceptual mechanisms that are assumed to be operative (right sub-column). L1 = activation of a triad; L2 = activation of a seventh chord; A1 = immediate anchoring; A2, A3, A4 = anchoring with respectively one, two or three intermediary tones; ? = unexplained percept.

<table>
<thead>
<tr>
<th>Stim. nr.</th>
<th>el. 1</th>
<th>el. 2</th>
<th>el. 3</th>
<th>el. 4</th>
<th>el. 5</th>
<th>3-percept</th>
<th>4-percept</th>
<th>5-percept</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F#</td>
<td>G</td>
<td>Bb</td>
<td>C</td>
<td>E</td>
<td>9</td>
<td>8</td>
<td>A1 6 A1</td>
</tr>
<tr>
<td>2</td>
<td>Bb</td>
<td>C</td>
<td>E</td>
<td>F#</td>
<td>G</td>
<td>11</td>
<td>L2 2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>F#</td>
<td>G</td>
<td>C</td>
<td>Bb</td>
<td>E</td>
<td>8</td>
<td>6</td>
<td>6 A1</td>
</tr>
</tbody>
</table>

On-line Processes
Stimulus presentation. Each series was preceded by a cadence consisting of the two chords C7 and F to induce the key of F-major. Stimuli were presented in a random order, different for each subject. Furthermore, each stimulus was presented at a different pitch height randomly chosen within a range of 6 semitones (the chords were transposed correspondingly). Inter-onset-intervals between the tones of the chords and the tones of the series were 800 ms, also the pause between the chords and the tones was 800 ms. The stimuli were generated by a Rhodes 770 synthesizer, using a Piano sound (Acoustical Piano

<table>
<thead>
<tr>
<th></th>
<th>G</th>
<th>C</th>
<th>Bb</th>
<th>E</th>
<th>F#</th>
<th></th>
<th>L2</th>
<th>10</th>
<th>L2</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>G</td>
<td>C</td>
<td>Bb</td>
<td>E</td>
<td>F#</td>
<td></td>
<td>L2</td>
<td>10</td>
<td>L2</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>G</td>
<td>C</td>
<td>E</td>
<td>F#</td>
<td>Bb</td>
<td></td>
<td>L2</td>
<td>10</td>
<td>L2</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>C</td>
<td>E</td>
<td>F#</td>
<td>Bb</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>E</td>
<td>F#</td>
<td>Bb</td>
<td>G</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>F#</td>
<td>Bb</td>
<td>G</td>
<td>C</td>
<td>E</td>
<td></td>
<td>A2</td>
<td>A2</td>
<td>A2</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Bb</td>
<td>G</td>
<td>C</td>
<td>E</td>
<td>F#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>E</td>
<td>F#</td>
<td>C</td>
<td>G</td>
<td>Bb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>F#</td>
<td>C</td>
<td>G</td>
<td>Bb</td>
<td>E</td>
<td></td>
<td>A1</td>
<td>A2</td>
<td>A2</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>C</td>
<td>G</td>
<td>Bb</td>
<td>E</td>
<td>F#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>G</td>
<td>Bb</td>
<td>E</td>
<td>F#</td>
<td>C</td>
<td></td>
<td>L2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>F#</td>
<td>C</td>
<td>G</td>
<td>Bb</td>
<td>E</td>
<td></td>
<td>A2</td>
<td>7</td>
<td>A2</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>C</td>
<td>G</td>
<td>E</td>
<td>Bb</td>
<td>F#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>G</td>
<td>E</td>
<td>Bb</td>
<td>F#</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>E</td>
<td>Bb</td>
<td>F#</td>
<td>C</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>C</td>
<td>G</td>
<td>E</td>
<td>F#</td>
<td>Bb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>G</td>
<td>E</td>
<td>F#</td>
<td>Bb</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>F#</td>
<td>Bb</td>
<td>C</td>
<td>G</td>
<td>E</td>
<td></td>
<td>A3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Bb</td>
<td>C</td>
<td>G</td>
<td>E</td>
<td>F#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>E</td>
<td>F#</td>
<td>C</td>
<td>Bb</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>C</td>
<td>Bb</td>
<td>G</td>
<td>E</td>
<td>F#</td>
<td></td>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Bb</td>
<td>G</td>
<td>E</td>
<td>F#</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>F#</td>
<td>C</td>
<td>G</td>
<td>E</td>
<td>Bb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>E</td>
<td>G</td>
<td>Bb</td>
<td>F#</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>G</td>
<td>Bb</td>
<td>F#</td>
<td>C</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>G</td>
<td>F#</td>
<td>Bb</td>
<td>C</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>F#</td>
<td>Bb</td>
<td>C</td>
<td>E</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Bb</td>
<td>C</td>
<td>E</td>
<td>G</td>
<td>F#</td>
<td></td>
<td>A4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>G</td>
<td>F#</td>
<td>C</td>
<td>Bb</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>Bb</td>
<td>E</td>
<td>G</td>
<td>F#</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>G</td>
<td>F#</td>
<td>C</td>
<td>Bb</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>F#</td>
<td>C</td>
<td>E</td>
<td>Bb</td>
<td>G</td>
<td></td>
<td>A4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>C</td>
<td>E</td>
<td>Bb</td>
<td>G</td>
<td>F#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>Bb</td>
<td>G</td>
<td>E</td>
<td>F#</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>C</td>
<td>F#</td>
<td>G</td>
<td>Bb</td>
<td>E</td>
<td></td>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>F#</td>
<td>G</td>
<td>Bb</td>
<td>E</td>
<td>C</td>
<td></td>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>C</td>
<td>F#</td>
<td>G</td>
<td>E</td>
<td>Bb</td>
<td></td>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>F#</td>
<td>G</td>
<td>E</td>
<td>Bb</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>C</td>
<td>Bb</td>
<td>F#</td>
<td>G</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>F#</td>
<td>G</td>
<td>E</td>
<td>C</td>
<td>Bb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>E</td>
<td>G</td>
<td>C</td>
<td>Bb</td>
<td>F#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>G</td>
<td>C</td>
<td>Bb</td>
<td>F#</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>C</td>
<td>Bb</td>
<td>F#</td>
<td>E</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>Bb</td>
<td>F#</td>
<td>E</td>
<td>G</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>G</td>
<td>E</td>
<td>C</td>
<td>F#</td>
<td>Bb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>Bb</td>
<td>F#</td>
<td>E</td>
<td>C</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>E</td>
<td>C</td>
<td>G</td>
<td>F#</td>
<td>Bb</td>
<td></td>
<td>A2</td>
<td>A2</td>
<td>A2</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>C</td>
<td>G</td>
<td>F#</td>
<td>Bb</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>G</td>
<td>F#</td>
<td>Bb</td>
<td>E</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>F#</td>
<td>Bb</td>
<td>E</td>
<td>C</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

On-line Processes
5) and played via a Kawai active speaker. Stimulus presentation was controlled by an Atari 1040 STf computer.

Procedure. Subjects were seated in front of the computer and used a button panel to activate the presentation of a stimulus. Pressing the button labeled '3' (all other buttons were deactivated) led to the presentation of the first three tones of the series. If desired, subjects could listen repeatedly to the fragment by pressing the same button until they had formed an opinion whether or not the fragment formed a musical percept. The concrete question asked was: "Could this series of tones be a (part of a) musical fragment?". They then responded by pushing the 'y' or 'n' key on the computer keyboard. If their answer had been 'yes', they pressed the '4' button to hear the initial 4 tones of the series which had to be dealt with in the same way as before, otherwise the next series was presented. If the 4-element fragment was also judged an acceptable musical fragment, subjects finally listened to the complete series by pressing the '5' button. In this self-paced way, which allowed subjects to stop occasionally to rest, responses on all 53 sequences were collected. The operation of the gating paradigm is schematically shown in Figure 1.

![Diagram of the gating paradigm](image)

Figure 1. The gating paradigm used in the experiment. 'Melody?' means: does the tone series lead to a musical percept?

Subjects. Twelve musically experienced subjects, graduate students and staff of the department of experimental psychology of the University of Nijmegen, participated in the experiment. All subjects were active music listeners and the majority also played a musical instrument for several years, although none had received any formal musical training. Median age of the subjects was 26 years.

Results

Overall the subjects did not find the task difficult to perform, although they commented that they would have preferred to have more response alternatives then just 'yes' and 'no'. When asked after the experiment what criterion they used to judge the tone series, the subjects often answered that they judged a tone series as forming a melodic percept when they could imagine a virtual continuation of the series. This shows that extrapolation (and consequently expectation) is an inherent aspect of the process of perception.

Because of the exploratory nature of the experiment, only a global and rather qualitative analyses was performed. For each of the initial fragments of sizes 3, 4, and 5 of all 53 series we determined the number of subjects that judged the fragment to be a melodic percept. The

On-line Processes
results are shown in the left sub-columns of the last three columns of Table 1. Note that the average number of yes responses decreases going from the column 3-percepts to the column 5-percepts because in the paradigm a longer fragment was not presented when a subject had answered no, i.e., that the fragment was non-melodic.

Subsequently we analyzed those fragments that the majority of subjects judged to be a melody. Because of the decreasing number of presentations for the longer fragments we used different criteria of inclusion for the three fragment sizes: 10 or more yes responses for the fragments of size 3; 8 or more for the fragments of size 4, and 5 or more for the fragments of length 5. An analysis of the structure of these fragments led to a few preliminary hypotheses as to the perceptual mechanisms the subjects may have used to arrive at their musical judgments. These hypotheses will be explained by means of a number of tone series that were used in the Experiment and that are shown in Figure 2. In the Figure, the fragment that according to the subjects formed a musical percept is indicated by black notes, while the remaining notes are displayed in gray.

Figure 2. Examples of tone series judged as leading to a musical percept. See text for explanation.

When presented with series 1 (referring to the numbers in front of the tone sequences, not to those above the clefs which designate the stimulus number in the experiment) subjects hear the three first tones as a musical percept. This percept can be understood as resulting from the activation of a triad chord, namely C major. Percept formation for series 2 and 3 can be understood as resulting from the activation of a seventh chord, namely C7. The perception of the first three sequences can thus be understood as resulting from the activation of a triad or a seventh chord, which are examples of labeling. In this paper the activation of a triad is called Labeling 1 (abbreviated L1) and the activation of a seventh chord Labeling 2 (L2).

The perception of series 4 and 5 cannot be understood on the basis of labeling alone. For these series it is hypothesized that the mechanism of anchoring plays a role in the percept.

On-line Processes
formation. Anchoring is based on the idea that there exists a hierarchical relationship between the tones in a key, with for instance the diatonic tones (the tones of the scale being higher in the hierarchy than the chromatic (or non-scalar) tones (Cooke, 1959, pp. 46-47; Lerdahl, 1988) and that tones that are lower in this hierarchy are attracted by tones higher in the hierarchy (Povel, 1996; Zuckerkandl, 1956). Bharucha (1984) defines anchoring as the mechanism by which a tone not belonging to the currently activated chord is linked to a tone of that chord. Phenomenally the anchored tone is perceived as an ornament of the tone to which it is anchored. A tone can only be anchored to a tone that is very close in pitch (1 or 2 semitones) and that follows the tone to be anchored, usually but not necessarily, the immediately succeeding tone. In series 4 the F# is anchored to the next neighbor G, while in series 5 there is an intermediate tone between the F# and G. In this paper anchoring to an immediately following tone is abbreviated as A1, anchoring with one intermediate tone as A2 etc. In Table 1 it can be observed that there are even some percepts of length 5 that might be explained by anchoring with two intermediate tones (stimuli 21 and 26) and with 3 intermediary tones (stimuli 30 and 35). By way of example, Figure 3 shows the hypothesized steps in the coding of the sequence F# G C Bb (note example 4 of Figure 2).

<table>
<thead>
<tr>
<th>Fragment size</th>
<th>Input</th>
<th>Process(es)</th>
<th>Percept</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>F# G C</td>
<td>Anchoring</td>
<td>F#[ G C ]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Labeling</td>
<td>{G C} \rightarrow fifth or C triad</td>
</tr>
<tr>
<td>4</td>
<td>F# G C Bb</td>
<td>Labeling</td>
<td>{G C Bb} \rightarrow C7</td>
</tr>
</tbody>
</table>

Figure 3. Hypothetical steps in the perceptual processing of the tone series F# G C Bb.

Next we examined how many of the fragments that were judged as leading to a melodic percept (on the basis of the criteria mentioned above) could be understood as resulting from the application of the hypothesized mechanisms. In Table 1, the supposed applied perceptual mechanisms for these fragments are indicated in the right sub-columns under the headings 3-percect, 4-percept and 5-percept. It can be seen that the 24 selected 3-percepts can be explained on the basis of the mechanisms L1, L2, A1 and A2 (note that A1 and A2 always implies the operation of the mechanism L1 or L2), except for stimulus number 32 (indicated with a question mark). All 14 selected 4-percepts can be understood as resulting from the application of the mechanisms L2, A1 and A2. The 18 selected 5-percepts can be understood as resulting from the application of the mechanisms A1, A2, A3 and A4, except for 3 stimuli (numbers 29, 32 and 44).

Discussion

This exploratory study has revealed a number of hypothetical perceptual mechanisms that listeners may have used to transform the elements in the presented tone series into a musical percept. In particular it has been found that two mechanisms: labeling (in fact one instance of labeling: the activation of a chord) and anchoring (the linking of a non-chord tone to a chord tone) can account for the majority of the formed percepts.

On-line Processes
Returning to the questions posed in the introduction to this experiment, we can state that subjects are indeed able to indicate whether or not a tone series forms a musical percept. However, whether a tone series forms a musical percept does not seem to be a yes-or-no characteristic, but a quality lying somewhere on a continuum with two poles: very clearly forming a musical percept and very clearly not forming a musical percept. Phenomenally, this quality seems to be related to the amount of processing (e.g., imagining possible continuations of the series) necessary to incorporate the tones into a meaningful perceptual organization.

The paradigm we have used in the study has proven to be quite useful, although not in all respects. Especially, the aspect that if an initial fragment is judged non-musical, longer fragments are not presented anymore, is a negative feature leading to an undesired reduction of the number of collected data for the longer fragments. When we designed the experiment we did not realize that a longer fragment of a tone series could lead to a musical percept, while a shorter fragment does not.

We may conclude that the study has been most instructive by suggesting a few preliminary perceptual mechanisms operative in the perception of tone series. Currently we are performing an experiment in which the proposed mechanisms are tested in a more formally designed experiment using an improved paradigm.

References


Acknowledgment

I thank Erik Jansen for running the experiment and for his help in analyzing the data.
Click Migration and Segmentation in Gaelic Melodies
Roisin L Ash
Department of Psychology
University of Stirling

Click localisation is a useful technique for examining our perceptual processing units within speech (Fodor & Bever, 1965; Ladefoged & Broadbent, 1960) and music (Stoffer, 1985). The basic task involves the listener attending to an auditory stimulus in which a superimposed click is positioned and subsequently indicating where the perceived click occurred. Results reveal that listeners tend to perceptually migrate the click towards the nearest phrase boundary. Basically, clicks positioned before a boundary will be perceived as occurring later than their correct position and clicks positioned after the phrase boundary will be heard as occurring before their actual position.

Click migration of music has been demonstrated in short artificial sequences of tones (Gregory, 1978) and rhythmically simple two-phrase melodies (Sloboda & Gregory, 1980), using a visual task in which listeners mark the perceived click on a notated score. Methodological criticisms raised by Stoffer (1985), include the confounding effect of visual segmentation of the sequences and the use of artificial melodies which lack ‘real’ phrase indicators. Other problems include the dichotic listening procedure which is predominantly used for click migration experiments. The click is presented to one ear while the musical stimuli is presented to the alternate ear. An effect of ear of presentation can occur using this method for both click perception and melody recognition (Kimura, 1964). Although some click localisation studies control for this possible confounding factor (Gregory, 1978; Sloboda & Gregory, 1980), others do not (Stoffer, 1985).

Many click migration studies have neglected to consider where the listener would naturally parse the experimental sequences. Rather, they have focused on relating any migration effects obtained to formal descriptions of the phrase structure of sequences presented to listeners. It may be that the phrase boundaries formally specified by trained musicians are the most likely perceptual segmentation points in a given sequence. Nevertheless, perceptual phrases should be experimentally defined. The most straightforward method for doing this is to ask the listener directly for their preferred segmentation of a sequence of events.

Experiments which have examined the listener’s preferred segmentation of music have typically employed some form of visual technique such as marking preferred segmentation positions a simplified notated score (Deliege, 1987). Others have used motor response task such as foot-pedalling (Clarke & Krumhansl, 1990) and button pressing (Krumhansl, 1996).

The present study used a button-press response to avoid potential visual cues from a visual representation of the score. Having established where the preferred segmentation points occur using this simple method (Experiment 1, Segmentation), a
modified click localisation technique was examined using these identified boundaries (Experiment 2: Click Migration). The click migration experiment differed from previous studies by using real melodies (unaccompanied Scottish Gaelic tunes) which were pre-recorded and presented in a binaural manner. The listener heard identical stimuli (melody excerpt containing an inserted click) in both ears at the same time. The experimental task avoided the visual segmentation effects by employing a non-visual motor and memory task (detailed below in the methods section).

Experiment 1: Segmentation
Method
Subjects The subjects were 24 undergraduates, 14 female, 10 male, aged between 18 and 36 years. Participation in the experiment was in part fulfilment of a psychology course requirement.
Stimuli/Apparatus The stimuli consisted of excerpts from three Scottish Gaelic melodies, s1, s2, and h1, shown in notated form in appendix 1. The melodies were selected from a large set that had been previously rated on a perceptual-emotional quality scale whereby listeners positioned a sliding pointer on a computerised visually displayed Happy - Sad scale. The melodies used in the present experiment were chosen from the extremes of the scale, two from the sad extreme and one from the happy extreme. These were produced on a NEXT using a bagpipe timbre and played at a tempo of one crotchet beat per 0.60 second without expressive timing and without dynamic variation. Listeners responses were detected by an electronic switch and recorded on a sound edit package on a Macintosh Centris 650.
Procedure Listeners were asked to segment the melody into phrases by pressing an electronic hand-held switch when they thought that a meaningful unit had occurred. They were instructed to listen to the melody excerpt three times without actively responding and then to practise segmenting the excerpt by depressing the switch. They were asked to practise until they were satisfied with their responses. Their responses were then recorded. Experimental trials lasted between 15 and 20 minutes.
Results Figure 1a, below, displays segmentation responses for melody S1. It was assumed that a meaningful unit would end on a note offset, thus responses are grouped into units of time related to note offsets. The response patterns for the three melodies used in the present experiment are summarised in Table 1, appendix 1.

Segmentation responses are fairly consistent for the three melodies with the majority of subjects using note duration as the major indicator of phrase boundary occurrence. As is evident from the example in Figure 1a, the majority of responses cluster around the longest notes in melody S1. As is shown in Table 1, appendix 2, this is characteristic of the responses to all three melodies. There are four main perceived boundaries in melodies S1 and S2 and three main boundaries in melody H2 which group around the longest held notes.
Figure 1a. Segmentation responses for melody s1. Responses were grouped into crotchet units of time (0.60 seconds, marked on the x-axis) in relation to note offsets. Percentage of subjects responding in a given 0.60 second period of time are shown on the y-axis while position in the melody in real time is shown on the x-axis. The dotted lines mark the onset - offset of the long notes in the melody. A quaver rest occurs between 13.2 - 13.5 seconds.

Experiment 2: Click migration

Method

Subjects The subjects were 69 psychology undergraduates, 48 females, 21 males, aged between 17 and 47 years. Participation in the experiment was in part fulfilment of a psychology course requirement.

Stimuli/Apparatus The stimuli consisted of excerpts of the first two phrases identified in each melody in the above segmentation experiment. The excerpts are described in Table 1, below.

<table>
<thead>
<tr>
<th>Melody</th>
<th>Duration (secs)</th>
<th>Number of notes</th>
<th>Number of clicks</th>
<th>Range of clicks</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>10.2</td>
<td>14</td>
<td>13</td>
<td>note 2 - note 14</td>
</tr>
<tr>
<td>S2</td>
<td>9.6</td>
<td>14</td>
<td>12</td>
<td>note 3 - note 14</td>
</tr>
<tr>
<td>H1</td>
<td>9.6</td>
<td>29</td>
<td>26</td>
<td>note 5 - note 29</td>
</tr>
</tbody>
</table>

The excerpts were produced in exactly the same way as Experiment 1. The additional click sounds used in the present experiment consisted of a 50 msec amplified noise burst positioned on note onsets. Each excerpt containing a click was recorded onto audio cassette tape and played on a SONY F442E stereo cassette deck at comfortable sound listening levels. This resulted in 13 excerpts for S1, 12 excerpts for S2 and 26 excerpts for H1. These differed in click position only and were identical in every other respect. Listeners responses were detected by an electronic switch and recorded on a sound edit package on a Macintosh Centris 650.

Procedure Listeners were split into three separate groups, performing with one melody excerpt only. Ten listeners performed on excerpt H1, ten on excerpt S2 and forty-nine on excerpt S1. They heard excerpt pairs in close succession. The first excerpt of each pair contained a click placed on a note onset. The first task was a motor detection task in which the listener was required to press a hand-held switch in response to hearing the click. The second task was a memory task in which the listener was required to remember where in the melody the click had been positioned. The second excerpt of each pair, a copy of the first excerpt pair without the click but
identical in every other respect, was then played. The listener was required to press
the switch at the position in which the click had occurred in the first excerpt. They
were specifically instructed to attend to the melody primarily, rather than the click, in
the motor detection task in order that they could perform the subsequent memory task
as accurately as possible. The experiment ran as follows:
1. Excerpt pair A: click is placed somewhere on a note onset. The listener performs a
motor detection task by pressing the switch when the click is perceived.
2. Excerpt pair B: excerpt A without the click. The listener performs a memory task
by pressing the switch where they think the click occurred in pair A. Practise trials
were run using excerpts from two similar gaelic melodies to enable the listener to
become accustomed to the task. Experimental trials lasted between 15 and 30
minutes.

Results
Data was grouped into before and after boundary positions for the three melodies by
taking the mean of listeners memory responses to click positions placed two beats
before a boundary and two beats after a boundary. Mean responses are displayed in
Figure 2, below and summarised in Table 2, appendix 3.

A series of repeated measures ANOVA'S performed on the grouped data revealed a
significant effect of pre and post-boundary click positions for excerpts S1 with
F(1,48) = 46.86, p < 0.001 and S2 with F(1,9) = 6.24, p < 0.05. Clicks placed before
these major boundaries were perceived as occurring an average of 0.20 seconds later
than their real position and clicks positioned after a boundary were perceived as
occurring an average of 0.52 seconds earlier than their real position. This significant
effect was not replicated for melody H1, although there is a migration trend for clicks
prior to the boundary perceived as occurring later than those after the boundary.

![Figure 2](image_url)

Figure 2 Migration effects for click positions before and after boundary 1.
The y-axis shows the mean memory responses of listeners. A score of zero indicates a correct response
(i.e. no difference between objective and perceived click position. Negative results are late responses
(the perceived click position occurs later than the objective position) and positive responses are early
responses (the perceived click position occurs before the objective click position).

Clicks placed before the second boundary in each of the melodies were perceived as
occurring an average of 0.54 seconds earlier than their actual position.
Discussion
The main point of interest in the segmentation experiment lay in identifying perceptual boundary positions, accounting for these in relation to the underlying structural features of the music and examining click migration effects at these boundaries using a novel non-visual method. The experiment revealed a powerful effect of note length on perceived boundary position within the three melodies, with longer notes very clearly allowing listeners to perceptually differentiate between groups in the melody. The melodies were segmented into three (H2) and four (S1 and S2) phrasal units. The temporal durations of the segmented units ranged from 3.3 seconds to 5.4 seconds in line with temporal limits offered by previous studies (Fraissee, 1982).

The present click migration technique revealed migration effects in the direction of the first phrase boundary. In line with previous research, clicks positioned directly before the boundary were perceived as occurring later and clicks positioned after the boundary were heard as occurring earlier than their actual position. A greater degree of click migration was found for clicks after the boundary than for clicks before the same boundary. This tendency for listeners to perceive the clicks as occurring earlier than their actual position is reflected in the earlier migration of clicks directly before boundary 2. Contrary to expectations, the direction of migration at these boundaries is away from the major boundaries. It is possible that temporal and attention factors, such as increased tendency to focus on the click as the melody progresses, are affecting the migration responses in these cases (Reber, 1970).

The present study shows that this is a useful non-visual technique to use. Clicks positioned within the first perceptual unit of a melody were migrated in the direction of the major phrase boundaries, as were clicks occurring directly after this boundary. Unexpected results of clicks away from second major boundary, highlight that we must consider the temporal positions of clicks in addition to their relation to a major boundary, especially when using this kind of memory task.

Appendix 1 The Gaelic Melodies used in Experiments I and II. Perceived phrase boundary positions are marked beneath the scores.

Melody S1

Melody S2

Melody H1
Appendix 2

Table 1 Duration of notes elicitng over 75% response in perceived boundary occurrence task and their temporal positions within the melody.

<table>
<thead>
<tr>
<th>Note duration</th>
<th>Time of note occurrence</th>
<th>% responses elicited</th>
<th>Time of note occurrence</th>
<th>% responses elicited</th>
<th>Time of note occurrence</th>
<th>% responses elicited</th>
</tr>
</thead>
<tbody>
<tr>
<td>quaver rest</td>
<td>13.2 - 13.5</td>
<td>79.2 %</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>dotted crotchet (0.90 seconds)</td>
<td>***</td>
<td>88.8 %</td>
<td>3.9 - 4.8</td>
<td>100 %</td>
<td>***</td>
<td>95.8 %</td>
</tr>
<tr>
<td>minim</td>
<td>3.6 - 4.8</td>
<td>94.4 %</td>
<td>8.7 - 9.6</td>
<td>100 %</td>
<td>***</td>
<td>8.7 - 9.6</td>
</tr>
<tr>
<td>dotted minim (1.80 seconds)</td>
<td>8.4 - 10.2</td>
<td>100 %</td>
<td></td>
<td></td>
<td>***</td>
<td>77.7 %</td>
</tr>
<tr>
<td></td>
<td>15.6 - 17.4</td>
<td>79.2 %</td>
<td></td>
<td></td>
<td></td>
<td>***</td>
</tr>
</tbody>
</table>

*** No notes of these values occurred within the melody

Appendix 3

Table 2 Mean memory responses to clicks positioned within two beats before (pre-boundary) and after (post-boundary) a boundary.

<table>
<thead>
<tr>
<th>Position of notes in melody</th>
<th>Melody S1</th>
<th>Melody S2</th>
<th>Melody H1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>st. dev.</td>
<td>mean</td>
</tr>
<tr>
<td>Pre - boundary 1</td>
<td>-0.309</td>
<td>0.706</td>
<td>-0.089</td>
</tr>
<tr>
<td>Post - boundary 1</td>
<td>0.531</td>
<td>0.732</td>
<td>0.518</td>
</tr>
<tr>
<td>Pre - boundary 2</td>
<td>0.830</td>
<td>0.722</td>
<td>0.416</td>
</tr>
</tbody>
</table>

References


A computer-aided Method of Analysis of melodic segmentation in monophonic Melodies

by Sven Ahlbäck, Royal University College of Music, Stockholm, Sweden and Dept. of Musicology, University of Gothenburg, Sweden

Background

Tonality and melodic segmentation

In the development of a computer-aided method for analysis of tonality in monophonic melodies of Swedish folk music, there was a need for a method of analysis of structural melodic segments. The term 'structural' refers to melodic segments that are revealed in the melodic structure as opposed to segmentation created in performance. The reason for this was the view that tonality perception in monophonic melodies is dependent on the perception of melodic segments, that tonality is perceived phrase-wise in such melodies. This assumption has been used by many folk music analysts, e.g. Béla Bartók and Zoltan Kodaly in their method of cataloguing Hungarian folk songs by the last note in every phrase. This view regards the phrase-level as the primary 'processing' level of conceiving a melody. Even if this process mostly is highly unconscious in terms of verbal expression, people often spontaneously are telling when this part of the conception of a melody doesn't work, in words like: "I don't understand this melody" or "it's just a lot of tones" etc.

The background of the project is an on-going study in musicology with the purpose to describe the tonal features of older Swedish folk music about tonality in other musical styles. Tonality is here used in its widest possible sense, meaning any system, perceived or structural, of relationships between pitches in music. This study has called for a method of analysing a large number of melodies. This was an important reason to try and make the method of analysis computer-aided.

There is an extensive analytical literature concerning phrase analysis, melodic segmentation and melodic grouping, but unfortunately mostly limited to Western art or popular music, with a major-minor tonal structure and dependent on functional harmony. Tonality is mostly regarded as a presupposition for analysis of melodic segmentation in these methods, while tonality analysis is the aim of this particular method. This method has therefor been developed independently from other methods of analysis of structural phrases, even though the result may show certain resemblance of other methods.

Because the idea is to study tonality in one musical style about tonality in other musical styles, the aim has been to develop a comparative method that is not entirely culture-dependent, but based on more general principles of melodic perception, to be able to study differences and similarities between different musical cultures, however with the limitation of monophonic melodies.

The limitations of the material

Since the raw material for this study basically consists of brief transcriptions of melodies in common notation the method of melodic segmentation have to be based entirely on information of relative duration, relative pitch and basic metrical patterns.

Earlier studies of the importance of different factors for listeners' segmentations of simple melodic structures seem to point in the direction that changes in acoustically prominent dimensions, as e.g. timbre or volume, are more important than changes in e.g. relative pitch.
(Bertrand 1994, Deliège 1987). With these results in mind, one can ask whether it's meaningful to study melodic segmentation from out of the limited information of brief transcriptions in common notation?

To answer that question I made some studies of melodic segmentation. But instead of using just simple melodic structures I used longer, more complex melodic examples, with identical pitch set, identical frequency of each of the pitches in the set and every note of equal duration. In the trials the subjects' task was twofold; (i) to segment the melody into groups at about beat level , (ii) to notate and sing the tonal center of the melody, if they could sense one. The instructions were given by example. The different melodic examples were constructed to give different responses concerning grouping (size and consistency) and tonal center. For one group of subjects there were no changes in articulation, volume and timbre in the examples, while there for the other group were periodical changes in these dimensions to test the strength of the different types of grouping factors. (Ex. fig 1)

The results showed almost no significant differences in segmentation at large or in perceived tonal center between the melodies with changes in articulation, volume and timbre and the same melodies without these changes. On the other hand the results showed clear and significant differences in grouping and perceived tonal center between the different melodies in the trials.

My interpretation of these results is that the acoustically weaker factors, e.g. change of melodic direction, seem to be more important for the listeners' interpretation of a complex melodic context than acoustically more prominent factors. The results also show that tonality perception is linked to perceived melodic grouping. This result is important since it indicates that the information content of brief transcriptions in common notation is sufficient for tonality analysis, at least on a basic level.

**The method**

**Basic principles**

I use melodic segmentation as a term for the perception of grouping of melodic events at any level of a melody, from beat groups to repeats. I am interested in the *perceivable* grouping and its structural counterparts, not just any principle of organisation of pitches or durations. Melodic segmentation can however be looked upon from two different angles, from the listeners point of view and from the creators/composers point of view, and sometimes I will use the second view since it can be useful to consider how melodic segments are created in a melody, rather than perceived, since a melody originally is created.

My method of analysis of melodic segmentation is based on the assumption that two primary structural principles are involved in the *creation* of melodic segmentation, perceived as well as constructed; (i) segments are created through repetition (the rule of similarity) and (ii) segments are created through discontinuance in structure (the rule of change). These two principles can in turn be regarded as a variation of the basic structural question: same or not same? From out of these two principles a set of methods based on analysis of rhythm, meter and pitch contour are formed, which are combined in a weighted analysis of melodic segmentation, which includes metrical structure and phrase structure at different levels.

The first of these principles refers to segmentation by repetition, something that is particularly common in instrumental music. The second principle is often especially important in vocal music from different cultures, and refers to the creation of melodic segments by structural interruptions, as for example rests, fermatas/long tones, shifts of register etc.

264
If we look at melodic segmentation from the listeners point of view it is useful to include a third principle, which can be regarded as a secondary grouping principle. This is the well-documented tendency to infer an established pattern on new events, of constancy, resistance against change. One example of this principle is metrical grouping, in which I include both the grouping in beats on different levels and in measures. Once a metrical pattern is established, we tend to interpret new events in relation to the metrical pattern; events that are on beats are perceived to be more articulated - foreground - while other events are perceived to be less articulated - background.

I am therefore asserting that phrases in metrical music are perceived about their position in the metrical structure. This is since the metrical structure, as pulse and grouping of beats in measures, is the basic reference of melody rhythm in such music.

If this is true, it follows that the analysis of melodic segmentation must be different in different metrical contexts.

**Three types of metrical organisation**

The repertoire of Swedish folk music can be classified into three different types of metrical organisation: symmetrical, asymmetrical and non-metrical contexts.

Symmetrical meter means that both beats and measures in general are of the same relative duration. In an asymmetrical structure are either beats or measures or both of different relative durations. In a non-metrical context there is neither beats, nor measures. (It is also possible to add a fourth category, quasi-metrical structure, where a metrical structure is interpreted as a non-metrical context, without general pulse)

The most evident example of melodies within Swedish folk music with non-metrical structure is herding calls, but other types of vocal music often belong to this category as well (or to the quasi-metrical category). The dance music and ceremonial music belong on the other hand almost entirely to the symmetrical or asymmetrical categories. Even if the symmetrical metrical structure is dominant, melodies with asymmetrical meter are not especially rare. In the western parts of Sweden are melodies with beats of different length common, while asymmetrical grouping at measure-level is more common in dance music from the eastern parts of the country.

**Grouping preference rules**

To be able to discuss how the method is interpreted in these different metrical contexts, we have to look at what aspects of melodic structure that I recognise in the analysis. On the basis of experiments (to which I will return below) I have construed a set of grouping preference rules, which can be said to reflect the listeners' preferences. I have obtained these rules from the results of the experiments and an elaboration of the basic principles mentioned above.

**Primary grouping principles**

I. The rule of similarity/proximity

Similar or close events tend to be grouped together.

- repetition of pitch, repetition of interval change or repetition of duration
- singular elements or arrays of elements which are similar or close according to pitch, melodic direction, interval size or duration creates grouping
- sequences, i.e. arrays of pitch- or/and duration-changes which are repeated, creates grouping

Sequences are dominant.
II. The rule of change/discontinuity

Change in melodic structure indicates grouping

The types of change are:

• change of melodic direction
• change of interval size
• change of pitch content
• change of tone/rest
• change of duration

Change of pitch content is overruled by change of melodic direction, which is overruled by change of interval size, which in turn is overruled by change of tone-rest and of duration

Secondary grouping principles

III. The rule of continuity (secondary grouping principle)

Grouping is inherited. Groupsize, start of group, etc. which is coherent with earlier grouping is preferred

• A sequence or repeated grouping creates an expectation of the same grouping to continue.

IV. The rule of symmetry (secondary grouping principle)

Symmetric grouping is preferred before asymmetric.

• Equal groups, e.g. according to duration, are preferred before asymmetrical (rather 2+2 than 3+1)
• groups which contain two (or multiples of two) elements are preferred before groups with three elements

Ternary grouping principles

V. The rule of articulation/gravity

An articulated (durationally or by pitch) element indicates grouping

• In sequences a more articulated element indicates start
• In non-sequential contexts a more articulated element indicates end
• start on tone is preferred before start on rest

VI. The rule of 'prägnanz'/contrast

Groups which contain a few contrasting elements is preferred before groups with no contrasting elements or many divergent elements

• Grouping with similar or close elements within the group, which are separated by contrasting elements are preferred before grouping where the contrast between elements within and outside the group are less contrast
• Groups with few elements are preferred before groups with many elements

It is important to realise that this is a brief summary of the rules, and that they hardly can be regarded as rules in a more narrow sense.

The dependency of metrical context

It is obvious that the strength of the 'rules' mentioned above is different in different metrical contexts: The presence of a pulse makes it possible for us to compare the length of phrases and thereby makes it possible to use the principle of symmetry and continuity, while the absence of a metrical structure makes the principles of change and gravity more important. While the symmetrical metrical structure can create an expectation of symmetry at higher levels, an
asymmetrical metrical structure makes us expect groups of different length and make it easier to change a grouping.

In non-metrical contexts where the rules of symmetry and continuity are irrelevant, the importance of rule of change is emphasised and vice versa - while rests or long tones in such contexts are powerful phrase boundaries it can be the starting point of phrases in symmetrical contexts.

While it takes quite strong means, e.g. a rhythmic and melodic sequence, to break a metrical pattern in a symmetrical context, it is not even recognised as asymmetrical in a non-metrical context.

I have therefore designed the method of analysis of melodic segmentation differently for the three different metrical contexts.

**A case study - 'strong' grouping preference**

**The experiment**

To study the relative importance of different factors for melodic segmentation in different metrical contexts I have made a series of tests with listeners.

Here I will focus on one of the experiments that concern the problem of under which conditions the expectancy of continuation of a symmetrical grouping is overruled by an asymmetrical grouping.

In this experiment the listeners, who were music students, had to choose among three different alternative groupings that represented three possible strategies, including both symmetrical and asymmetrical groupings. Then they had the possibility to make changes in the grouping in the selected alternative. (examples by fig. 2)

The results showed that even if the cultural musical background of the subjects is predominantly symmetrical, asymmetrical groupings were preferred by almost all subjects when the grouping indication was structurally evident (see below).

From these results I have obtained an order of priority between different structural factors for melodic segmentation. The results made it clear that I had to separate between 'strong' grouping indications, which overrules prior grouping expectations, and 'weak' grouping indications, which are dependent on the metrical context in which they appear.

**Conditions for 'strong' contour sequences**

From the results of the case study I obtained the following conditions for a strong contour sequence:

a) The groups have to contain at least three elements.

b) The melodic direction must be identical within the groups.

c) The interval relationship (at large) must be preserved between the groups, i.e. if the first note in the second group is higher than the first note in the first group all correspondent notes have to have the same relationship.

d) Within the sequence the groups have to be discrete in the sense that the interval, pitches or the change of direction between the groups and either before or after the sequence, are unique in relation to the intervals, pitches or changes of direction within the groups.
e) The sequence has to be discrete in the sense that at least two of the changes that constitute the groups of the sequence have to be different from each other.

f) The sequence must not be overlapped or 'disguised' by other strong grouping indications, as overlapping sequences, melodic return, repeated tones, etc.

These and other sets of rules are implemented in the computer programme, and tested in different contexts. As can be seen from the present example of asymmetrical metrical structure (ex./fig. 3), it gives the same results as the main result of the test with listeners.

In a melody with symmetrical metrical structure the results will be quite different. (ex. fig. 4)

**The generality of the rules**

My studies with listeners indicate that certain groupings are strongly preferred, but there is always a possibility to experience the grouping differently, especially in more ambiguous melodic situations.

The way the rule system is implemented in my programme it will always choose one solution before other solutions, and so it will always be some people who in some situations would experience the grouping differently from what the computer suggests. This problem is however in reality not that serious as one could come to believe if we look at a certain melodic situation, since there in real music mostly is extra evidence, redundant information, which helps to make the grouping evident. It is quite understandable that melodies contain redundant information if we conceive a melody as a means of communication.

But how culturally limited is this method? To study this I used music from different cultures, e.g. some examples of south Indian music where I had access to the melodic segmentation made by a distinguished performer of that kind of music. The result obtained by the computer analysis coincided to a degree of about 95% with the groupings preferred by the performer. (ex./fig. 5)

This is not to be interpreted as evidence of the melodic grouping principles implemented in this computer programme are universal, but it suggests that the rules are perceptually relevant in two different musical cultures to a degree that the programme can be useful for studying cultural differences in melodic grouping and tonality.

References


(additional figures will be presented at the conference)
THE TRANSCRIPTION OF PITCHES IN TONAL MELODIES.
A comparative study.

ISABEL CECILIA MARTINEZ
Universidad Nacional de La Plata - Argentina

Introduction
Listening to music is a process that -according to Sloboda- “involves forming an abstract or symbolic internal representation of the music”(1985). This representation has to be inferred from a variety of musical responses that may range from verbal descriptions to the encoding and writing of a musical piece.

A tonal melody can be perceived as a succession of pitches enchained according to an internal logic. Thus, pitches may be linked into syntactic micro-units, which we will called melodic patterns. The melody -understood as a complete sense unit- is a succession of patterns that form a syntactic chain. So as to memorize the pitches of a tonal melody and write them by dictation, the listener must solve the melodic chain by comparing the pitches within the same pattern and between different patterns. Previous studies (Martinez 1996, a and b; 1997, a) have shown that: i) the mere resolution of melodic contour is not enough to achieve the transcription of the melody; ii) the processing of interval relationships is different whether the pitch interval is formed within the same pitch pattern or between connective pitch patterns; and iii) the wrong measurement of the direction and/or size of an interval in the concatenation of pitches may lead to a following succession of slided pitches, and consequently derive in a sliding melodic contour.

We have considered to what extent it is likely that the listener may encode the tonal pitches in a hierarchical way, in order to make an in-depth analysis of the previous results. We have also assumed that this hierarchical encoding of tonal pitches follows the tonal hierarchy of the given melody. In order to achieve the transcription of a melody, the hierarchical structure might be of importance in the processing of pitch information during listening and later memorization. Thus, three variables were identified as probable indicators of the results of the transcription: i) pitch structural groupings; ii) pitch sequential relationships; and iii) metrical considerations for both i) and ii).

An experiment was designed to test the ability to transcribe pitches (post memorization) of melodic isolated patterns and to transcribe pitches of those very same patterns forming a complete tonal melody. In the study of the experimental melody we applied both the model of internal representation of pitch sequences in tonal music developed by Deutsch and Feroe (1981) and some peculiarities of the Schenkerian analysis as analytical tools. The results of this analysis were then contrasted with the results obtained in the study of fourteen subjects (N=14) who achieved 100% of correct results in the test of the transcription of patterns, but who made mistakes in the test of the transcription of the melody. If -following Deutsch (1981)- the musical structure reflects the ways in which musical information is mentally processed, then the three identified variables may be considered as indicators which are liable to explain the obtained results.

Theoretical Foundations
The complex nature of music requires that, while listening to a melody, the listener should process the information that has just listened to, and that he/she should simultaneously pay attention to the incoming string of information. Because of the limited capacity of short-term
memory, it is assumed that he/she must process the successive events in an organised way, and
that this process allows him/her to acquire a mental representation.

The grouping condition of pitches works on the basis of simple perceptual mechanisms. The
organisational principles that govern this process (for example the principle of “good
continuation”) derive from the Gestalt theory, which states that several notes are perceived to
be linked together. Concerning this principle, several organisational factors that may create
grouping conditions for pitch perception have been studied, such as the temporal proximity
between events (Deutsch, 1982) and the identity of melodic contour (Dowling, 1978).

As far as tonality is concerned, tonal music is regarded as a pitch hierarchy in which more
complex mechanisms of representation are involved. From the point of view of the
Schenkerian approach, three structural levels can be basically distinguished: at the lower level
there is the foreground, or melodic surface representation; at a higher level there is the middle
ground, and at the highest level there is the background, where harmonic structure is
abstracted. Several structural descriptions of the perception of a piece have been proposed,
such as Lerdhal and Jackendoff’s (1983) model, from which studies have been derived that
are related to the perception of underlying structures (Bigand, 1990). Concerning structural
descriptions of the surface, several factors have been studied (for example, temporal
segmentation in connection with pitch sequences and tonal structure, Deutsch, 1980).

Deutsch and Feroe (1981) developed a model to explain the relationships between the
hierarchical organisation of tonal music and the perception of pitch sequences. According to
this model, tonal hierarchy favours pitch retention and processing while listening. Their model
is characterized as an architectural network, in which pitch sequences are represented at
different levels. Thus, pitch elements which form structural units, and which are present at any
given level, are elaborated by further pitch elements at a lower level, forming pitch
subsequences. These pitch subsequences are, in turn, pitch structural units represented at the
lower level. This process goes on until the melodic surface is reached. The redundant pitch
information that provides the multiple representations should make it easier to retrieve this
information from memory.

Deutsch and Feroe’s model presents a specific notation or language to describe serial pitch
patterns, and it is assumed that this specific language reflects specific encodings during pitch
processing. The basic elements of this notation are: i) pitch alphabets (a, β, etc.) - linearly
ordered sets of symbols, such as the diatonic scale; ii) elementary operators, referred to a pitch
alphabet (e.g. next (n), same (s), and previous (p), which provide order to the pitch
succession); iii) structures(A,B,C, etc.) -successions of elementary operators--; iv) sequences of
notes (S) - structural units of pitch information interpreted as an organized set of elements
which belong to an alphabet, which have a specific structure, and which refer to one of the
notes: the reference element (a dominant pitch considered a structural goal); and v) sequence
operators which combine two or more sequences to form compound sequences. The basic
sequence operator is called Prime (pr): the pitch material here is presented without any
alteration. The other two sequence operators are Retrograde (ret) Inversion (inv). Ret refers to
the presentation of the original material in reverse order and inv refers to the presentation of
the sequence of notes in such a way that ascending intervals are presented as descending, and
vice versa. Both ret and inv are defined as elaborations of pr.

The Gestalt principles of proximity and good continuation govern the pitch organization at
each hierarchical level. Therefore, the configuration of a melody results from mental
operations in which a succession of pitches is represented as structural units at the surface and, at the same time, as structural units at higher levels, units in which some events acquire relevance over others. The representation of pitches at higher levels brings support to the pitch representation at lower levels, enabling the subject to retrieve the pitch chain at the surface level.

In the perception of a piece of music, in a real listening situation the listener is frequently faced with the complexity of musical material. If perceptual mechanisms lie at the most basic level of pitch representation, therefore the most univocal representation would be that which presents no conflict between the structural hierarchy of the piece and the perceptual grouping mechanisms. Temporal segmentations of pitch groupings that match the tonal hierarchies will favour a correct processing, thus fostering the retrieval of the pitch information from memory. On the contrary, if temporal grouping is in conflict with the tonal hierarchy of the stimulus, the unstructured segmentation can be an obstacle to form a correct representation. Thus, the representation of a piece of music while listening is mediated by conditions of more or less important ambiguity in the characteristics of the musical material. The present study claims that the structural groupings, the sequential incidences, and the metric position of the groupings are appropriate indicators which can explain the results obtained in the pitch transcription (post memorization) of a tonal melody. The study further aims at identifying the way in which the interaction of the tonal, sequential, and metrical factors may lead to generate certain conditions of structural ambiguity and to have incidence in the obtained results.

Methodology

An experimental test was designed, and it was administered in two sessions under two experimental conditions:

1: Audition, memorization and transcription by dictation of the pitches of eight melodic patterns. The musical stimulus had the following characteristics: melodic contour that proceeded by steps; rhythm in ternary division; harmonic accompaniment. Four of these melodic patterns were considered the original patterns (drawn from the complete melody, presented in identical rhythmic and harmonic- tonal situation with respect to the melody) and the remaining four, were considered distractors patterns (identical to the originals patterns in the melodic-rhythmic design, but presented in a different tonality and with different harmonic chords). The patterns were presented on a recording. Each pattern was repeated three times, with a one-minute period between each of the eight examples. During this one period the pitch material which had been memorized was asked to be transcribed. The examples were presented in random order, and the first note of each pattern was provided so as to accomplish the transcription on the stave.

2: Audition, memorization and transcription by dictation of a tonal melody of the following characteristics: thirty pitches by steps in the ambit of an octave; sixteen beats; ternary division; key of C Major; central register; and moderate tempo. The melody is confonned by the four original patterns of experimental condition 1. The melody was presented complete and with harmonic accompaniment. The fist note was provided so as to accomplish the transcription on the staves. The number of repetitions was rated at four, with an interval of time equivalent to the length of the melody between presentation and repetitions for retention. (The interval of time was given previous to the transcription). Ten minutes were allotted for transcription. There was a one-week period between one test session and the next.

Sample: The subjects (N=14) were first-year students from the music courses at the Facultad de Bellas Artes (UNLP). The sample was conformed with those subjects who had obtained 100% of correct results in the transcription of the isolated patterns (session 1), and who nevertheless presented mistakes in the transcription of the complete melody (session 2).
Results

Deutsch and Feroe's model was applied to the analysis of the experimental melody, and the result of such analysis was as follows:

**Experimental melody**

```
I   V, I I IV V 7

Tree Diagram
```

```
G
G
```

A= (n4, *)
B= (*, 2 s)
C= (3 n, *)
D= (2 p, *)
E= (s, n, *)

**Resultant Notation**

\[ S=A[p_r](B,C) [p_r] ([*],D,E, 4 {*}) ] [ 2p_r,2r_c,p_r,r_c,5p_r] (F, 3G,2H,{*}), F, I, 2{*}), c: C. 

(where C is the C Major diatonic scale and c is the reference element)

Structure is defined in four levels corresponding levels 1 and 2 to the background, level 3 to the middleground, and level 4 to the foreground or surface. The melody is divided into two parts: the first part prolongs g, being the second part a progression towards c.

A comparative analysis of results (N=14) was done in structural levels.

<table>
<thead>
<tr>
<th>level</th>
<th>description</th>
<th>resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A= (n4, *) = g, c.</td>
<td>9</td>
</tr>
<tr>
<td>2 (1st part)</td>
<td>B= (*, 2 s)= g, g, g</td>
<td>8</td>
</tr>
<tr>
<td>2 (2nd part)</td>
<td>C= (3 n, *)= f, e, d, c</td>
<td>5</td>
</tr>
<tr>
<td>3 (1st part)</td>
<td>D= (2 p, *)= c, f, g; E= (s, n, *)= b, s, g</td>
<td>8 (1)</td>
</tr>
<tr>
<td>4 (1st part)</td>
<td>F=(*, n, p)=g, a, g; G=(2n, p *)=f, c, d, c; ret G=e, d, c, f; ret G=f, e, f, g;</td>
<td>7 (2)</td>
</tr>
<tr>
<td>4(2nd part)</td>
<td>H=(<em>, 2n)=a, b, c; ret H=c, b, a; (</em>)=g</td>
<td>2 (3)</td>
</tr>
</tbody>
</table>

(1) out of 8 subjects who correctly solved level 2, 8 correctly solved level 3
(2) out of 5 subjects who correctly solved levels 2 and 3 (1st part), 7 correctly solved the surface level (level 4; 1st part)
(3) out of 5 subjects who correctly solved level 2 (2nd part), only 2 subjects solved the surface level (level 4; 2nd part)
The prolongation denotes a superior level of resolution compared to the progression. So as to estimate the significance of that difference, the number of correct responses for each pitch was calculated. The mean of resolution of each part was as follows: for the prolongation 11.41; for the progression 7.5 (t=4.789 p<.001).

**Interpretation and Discussion**

The model of the internal representation of pitch sequences in tonal music developed by Deutsch and Feroe (1981) has turned out to be a useful instrument to describe the performance of 14 subjects in the pitch transcription of a tonal melody. It is useful because it allows an organisational description of the melodic material in terms of sequences and relationships between sequences in a way that may reflect the complex nature of the representation of the tonal hierarchy.

The number of subjects who correctly transcribed the structural pitches (reference elements) of the melody was higher to the number of subjects who correctly transcribed the rest of the pitches. This suggests that structural pitches are configured as a frame which may support the pitch processing at lower levels. Nevertheless, the results show that -although the auditor may well configure the higher structural levels (background and middleground)- the accidents that occur at the surface level reveal that the relationships between the surface level and the higher levels require further analysis. As it was noted above, the performance of the subjects evidenced a clear difference between the two parts of the melody. Some characteristics of the configuration of the melody illustrated by the analysis may explain the differences found.

If **structural groupings** are considered, we have the following result: in the **1st part** (prolongation) we can identify three structural levels -levels 1 and 2 correspond to the background, level 3 correspond to the middleground, and level 4 correspond to the foreground-. **Each lower level is an elaboration of the higher level.** On the contrary, in the **2nd part** (progression ) we find the structural line at level 2, and the surface level at level 4. There is no elaboration at level 3 (middleground). The identified difference assigns an addtional structural value to the first part of the melody. We can assume that the elaboration at the middleground level may be a frame which simultaneously supports the processing at higher and at lower levels.

If **sequencial elements** are considered, if we focus on the first part of the melody, the analysis shows that only one structural unit \(G=(2n,p,*)\), two sequence operators \((pr \text{ and } ret)\), and three operations \((pr, ret, ret)\) are required to solve the encoding of twelve out of the total number of pitches of the first part. The sequence operator \(ret\) reduces the number of structural units required to form the correct representation, making it possible to obtain a considerable reduction in the pitch information retrieve from memory.

If **metrical incidences** are considered, the analysis shows the following results: in the first part of the melody the first two structural units form a melodic contour which corresponds to this succession of notes: \(g, a, g, f, e, d, e\). In the second part of the melody there appears this melodic contour which is almost identical to the referred sequence, but which presents a different harmonic and metrical position with the addition of one note. Thus we have: \(f, g, f, e, d, c, d, c\). According to Dowling (1978) if the melodic contour is presented several times, the auditor will tend to form groupings on the basis of the identity of this melodic contour. Now, if the auditor has just configured a specific melodic contour representation, when faced with a new pitch sequence that is similar in contour but different in metrical and harmonic-tonal
position, it is likely that an alternative representation would be realized. This would create a conflict in the configuration of the melodic surface, for both representations might be present when retrieving the information from memory. Deutsch’s analysis shows that, in the second part, the representation of the succession of pitches requires four structural units: \( F = (\ast, n, p); I = (\ast, 2p); \ast, \ast \), and four sequence operators (pr, pr, pr, pr), while the initial referred fragment requires only two structural units and two sequence operators.

If the harmonic-metrical structure is analyzed, it will be observed that in the first part of the melody, the metric accent appears every two beats, while in the second part, the harmonic rhythm changes and the metric accent occurs beat by beat. This generates the segmentation of the surface in an increasing number of melodic structural units, each with a fewer number of pitches. If we take into account the fact that, out of the five subjects that configured the structure of the second part, only two of them solved the surface level, we may describe this area as the more ambiguous regarding the structural hierarchy. Incidentally, this area obtained the lowest level of resolution, both at the surface level and at the background level. Nevertheless, in the second part, the two last notes of the melody (d, c) are precisely the notes which obtained the highest level of resolution. This goes on to prove the high structural value of these two elements. In the analysis, each of them is represented as a structural unit of a single element (\( \ast \)).

As these “accidents” occurred in the last part of the melody, the study of structural ambiguity would be the aim of further research. This study should focus on locating the ambiguous segment at the beginning or in the middle of the melody.

Paraphrasing West, Howell and Cross (1991) in the field of musical perception there are “big questions to which at present we can only offer partial and inadequate answers”...

References


Pitch Direction and Melodic Expectancy:  
A Reexamination of Empirical Data

Zohar Eitan, Department of Musicology, Tel Aviv University*

The Implication-Realization (I-R) Model (Narmour 1990, 1992) proposes that melodic structure and its cognition are shaped by implicative events, generating specific expectations, and by the consequent realization or denial of these implications. Though acknowledging that pitch ascent and descent differ in their musical import (Narmour 1990: 362-64), the model is basically symmetrical with regard to pitch direction: a melodic interval or pattern and its pitch inversion (e.g., C4-D4-A4 and A4-G4-C4) are similarly categorized and generate similar implications. Thus, the model's two basic principles of melodic implication relate the magnitude of an implicative melodic interval to the magnitude and direction of the realized interval following it, regardless of the implicative interval's direction: the first principle, Intervallic Difference, maintains that small pitch intervals (1-4 semitones) of either direction imply intervals similar to them in size, while large ones (larger than 7 semitones) imply intervals smaller in size. The second, Registral Direction, maintains that small intervals—both ascending and descending—imply continuity of direction, while large ones imply directional change.

Several studies of melodic expectancy have recently tested the psychological validity of the I-R model. This paper examines the effect of pitch direction on melodic expectancy by reanalyzing data of four of these studies: three experiments conducted by Krumhansl and Schellenberg (Krumhansl 1995, Schellenberg 1996), and one by Cuddy & Lunney (1995). These analyses provide a basis for examining the functions of pitch direction in melodic expectancy, and in particular its roles in an implication-realization model of melody.

Reanalyzing Expectancy Studies

In Cuddy and Lunney (1995) subjects were presented with eight melodic intervals, two “small” intervals (a Major Second and a Minor Third), and two “large” ones (a Major Sixth and a Minor Seventh), each in both directions. Each of these intervals was continued, in separate trials, by 25 test tones, comprising all chromatic tones ranging from an octave below to an octave above its second note. Subjects were asked to rate numerically...

* I am grateful to Lola Cuddy and Glen Schellenberg for providing their data for reanalysis, to Eugene Narmour for valuable comments, and to David Steinberg for his help in statistical analysis.

1. Similarity of interval magnitude is defined in the I-R model as follows: when no change of direction occurs, intervals differing by 3 semitones or less are considered similar; when direction changes, the similarity threshold is reduced to 2 semitones.

2. Medium, “threshold” intervals (5-7 semitones) may imply either continuation or reversal of interval size and direction.
(on a scale from 1 to 7) how well each test tone continued the initial (implicative) interval. In Schellenberg and Krumhansl's three experiments, actual melodies (from British folksongs, posttonal Lieder by Webern, and Chinese folksongs) were used as stimuli. These melodies were interrupted, and the interrupted tone was followed by a test tone (test tones again comprised all chromatic tones ranging from an octave below the last tone of the interrupted melody to an octave above it). As in Cuddy and Lunney's experiment, subjects were asked to rate numerically how well the test tone continued the melody.

Table 1: Compared melodic patterns.

Pairs 1-6 examine the effect of the implicative intervals’ direction on the direction or magnitude of realized intervals. Pairs 7-12 examine combinations of realized magnitude and direction, defining the I-R model’s Basic Melodic Structures. Each of the pairs 1-12 was examined (a) considering only small implicative intervals (b) considering only large implicative intervals, and also considering all implicative intervals (average of a and b).

<table>
<thead>
<tr>
<th>Examined Structure</th>
<th>Direction of Realized Interval</th>
<th>Magnitude of Realized Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Continuity of Direction</td>
<td>same</td>
<td>all</td>
</tr>
<tr>
<td>2. Change of Direction</td>
<td>different</td>
<td>all</td>
</tr>
<tr>
<td>3. Similarity of magnitude</td>
<td>all</td>
<td>similar</td>
</tr>
<tr>
<td>4. Increase of magnitude</td>
<td>all</td>
<td>increased</td>
</tr>
<tr>
<td>5. Decrease of magnitude</td>
<td>all</td>
<td>decreased</td>
</tr>
<tr>
<td>6. Change of magnitude</td>
<td>same</td>
<td>increased or decreased</td>
</tr>
<tr>
<td>7. Process (P)</td>
<td>different</td>
<td>similar</td>
</tr>
<tr>
<td>8. Registral Return (IP/ID)</td>
<td>same</td>
<td>increased</td>
</tr>
<tr>
<td>9. Registral Process (VP)</td>
<td>different</td>
<td>decreased</td>
</tr>
<tr>
<td>10. Reversal (R)</td>
<td>same</td>
<td>increased</td>
</tr>
<tr>
<td>11. Intervallic Reversal (IR)</td>
<td>different</td>
<td>all</td>
</tr>
<tr>
<td>12. Registral Reversal (VR)</td>
<td>same</td>
<td>small</td>
</tr>
<tr>
<td>13. Realized intervals (all)</td>
<td>different</td>
<td>medium</td>
</tr>
<tr>
<td>14. Realized intervals (small)</td>
<td>similar</td>
<td>large</td>
</tr>
<tr>
<td>15. Realized intervals (medium)</td>
<td>increasing</td>
<td></td>
</tr>
<tr>
<td>16. Realized intervals (large)</td>
<td>decreasing</td>
<td></td>
</tr>
</tbody>
</table>

My reanalyses compared average ratings for melodic structures initiated by ascending intervals with those of equivalent structures initiated by descending ones. A melodic structure in this study is a combination of three melodic notes, comprising an implicative interval (the interval between notes 1 and 2) and a realized interval (between notes 2 and 3). Three parameters distinguishing melodic structures from each other are considered: the direction and magnitude of the implicative interval, and the direction and magnitude of the realized one. The direction of implicative intervals is ascending or descending (data for oblique implicative intervals were excluded from the analysis). The magnitude of implicative intervals is categorized, following the I-R model’s definitions, as either small or large. The direction and magnitude of realized intervals are defined relatively to those of the implicative intervals preceding them: direction as same or different, magnitude as decreasing, similar, or increasing (similarity of magnitude is defined in accord with the I-R model; see footnote 1).

3. In Schellenberg/Krumhansl the implicative interval is the last interval of the interrupted melody.
Table 2. Average Scores

Rows 1a, 2a, etc. present scores for small implicative intervals; in rows 1b, 2b, etc. scores are for large implicative intervals; rows 1, 2, etc. present average scores for implicative intervals of all sizes. Asterisks indicate the statistical significance of the difference between average scores of the "ascending" and "descending" member of each pair: * signifies p<.05; ** signifies p<.01; *** signifies p<.001.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>CUDDY &amp; LUNNEY</th>
<th>SCHELLENBERG (BRIT.)</th>
<th>SCHELLENBERG (WEB.)</th>
<th>SCHELL. (CHIN.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Asc</td>
<td>Desc</td>
<td>Asc</td>
<td>Desc</td>
</tr>
<tr>
<td>1. Dir. cont.</td>
<td>4.22</td>
<td>3.62***</td>
<td>3.43</td>
<td>2.93***</td>
</tr>
<tr>
<td>1a. (Sml. imp.)</td>
<td>4.26</td>
<td>3.75***</td>
<td>3.63</td>
<td>3.18**</td>
</tr>
<tr>
<td>1b. (Lrg. imp.)</td>
<td>4.18</td>
<td>3.49***</td>
<td>3.24</td>
<td>2.67***</td>
</tr>
<tr>
<td>2. Dir. chng.</td>
<td>4.19</td>
<td>4.63***</td>
<td>4.25</td>
<td>4.34</td>
</tr>
<tr>
<td>2a. (Sml. imp.)</td>
<td>4.03</td>
<td>4.45**</td>
<td>4.05</td>
<td>3.89</td>
</tr>
<tr>
<td>2b. (Lrg. imp.)</td>
<td>4.33</td>
<td>4.22</td>
<td>4.11</td>
<td>3.96</td>
</tr>
<tr>
<td>3. Mgn. Sim.</td>
<td>4.33</td>
<td>4.78</td>
<td>5.18</td>
<td>4.84*</td>
</tr>
<tr>
<td>3a. (Sml. imp.)</td>
<td>3.61</td>
<td>3.59</td>
<td>2.84</td>
<td>2.52</td>
</tr>
<tr>
<td>3b. (Lrg. imp.)</td>
<td>3.33</td>
<td>5.04**</td>
<td>2.62</td>
<td>4.23***</td>
</tr>
<tr>
<td>4. Mgn. Inc.</td>
<td>3.60</td>
<td>3.64</td>
<td>2.83</td>
<td>2.62</td>
</tr>
<tr>
<td>4a. (Sml. imp.)</td>
<td>3.61</td>
<td>3.59</td>
<td>2.84</td>
<td>2.52</td>
</tr>
<tr>
<td>4b. (Lrg. imp.)</td>
<td>3.33</td>
<td>5.04**</td>
<td>2.62</td>
<td>4.23***</td>
</tr>
<tr>
<td>5. Mgn. Dec.</td>
<td>4.64</td>
<td>4.55</td>
<td>4.63</td>
<td>4.42*</td>
</tr>
<tr>
<td>5a. (Sml. imp.)</td>
<td>5.33</td>
<td>5.00</td>
<td>5.15</td>
<td>5.81</td>
</tr>
<tr>
<td>5b. (Lrg. imp.)</td>
<td>4.61</td>
<td>4.54</td>
<td>4.60</td>
<td>4.32**</td>
</tr>
<tr>
<td>6. Mgn. Chng</td>
<td>4.10</td>
<td>4.08</td>
<td>3.68</td>
<td>3.46*</td>
</tr>
<tr>
<td>6a. (Sml. imp.)</td>
<td>3.67</td>
<td>3.64</td>
<td>2.97</td>
<td>2.70</td>
</tr>
<tr>
<td>6b. (Lrg. imp.)</td>
<td>4.57</td>
<td>4.39</td>
<td>4.47</td>
<td>4.32</td>
</tr>
<tr>
<td>7. P</td>
<td>4.31</td>
<td>3.69***</td>
<td>3.58</td>
<td>3.31***</td>
</tr>
<tr>
<td>7a. (Sml. imp.)</td>
<td>4.91</td>
<td>4.62*</td>
<td>5.06</td>
<td>4.51***</td>
</tr>
<tr>
<td>7b. (Lrg. imp.)</td>
<td>3.80</td>
<td>2.91***</td>
<td>2.46</td>
<td>2.09*</td>
</tr>
<tr>
<td>8. IP/ID</td>
<td>4.36</td>
<td>4.86***</td>
<td>4.74</td>
<td>4.94</td>
</tr>
<tr>
<td>8a. (Sml. imp.)</td>
<td>4.68</td>
<td>4.94*</td>
<td>5.29</td>
<td>5.18</td>
</tr>
<tr>
<td>8b. (Lrg. imp.)</td>
<td>4.40</td>
<td>4.77***</td>
<td>4.18</td>
<td>4.70**</td>
</tr>
<tr>
<td>9. VP</td>
<td>3.71</td>
<td>3.02**</td>
<td>2.55</td>
<td>2.19*</td>
</tr>
<tr>
<td>9a. (Sml. imp.)</td>
<td>3.71</td>
<td>3.02**</td>
<td>2.55</td>
<td>2.19*</td>
</tr>
<tr>
<td>9b. (Lrg. imp.)</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>10. R</td>
<td>4.66</td>
<td>4.82</td>
<td>4.85</td>
<td>5.00</td>
</tr>
<tr>
<td>10a. (Sml. imp.)</td>
<td>3.33</td>
<td>5.00</td>
<td>5.15</td>
<td>5.81</td>
</tr>
<tr>
<td>10b. (Lrg. imp.)</td>
<td>4.61</td>
<td>4.81</td>
<td>4.82</td>
<td>4.91</td>
</tr>
<tr>
<td>11. IR</td>
<td>4.62</td>
<td>4.17**</td>
<td>4.27</td>
<td>3.44***</td>
</tr>
<tr>
<td>11a. (Sml. imp.)</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>11b. (Lrg. imp.)</td>
<td>4.62</td>
<td>4.17**</td>
<td>4.27</td>
<td>3.44***</td>
</tr>
<tr>
<td>12. VR</td>
<td>3.51</td>
<td>4.14**</td>
<td>3.05</td>
<td>2.96</td>
</tr>
<tr>
<td>12a. (Sml. imp.)</td>
<td>3.52</td>
<td>4.08**</td>
<td>3.10</td>
<td>2.82</td>
</tr>
<tr>
<td>12b. (Lrg. imp.)</td>
<td>3.33</td>
<td>5.04**</td>
<td>2.62</td>
<td>4.23***</td>
</tr>
<tr>
<td>13. Realized</td>
<td>4.40</td>
<td>3.86***</td>
<td>3.82</td>
<td>3.51**</td>
</tr>
<tr>
<td>14. Sm. Rlz.</td>
<td>4.81</td>
<td>4.69</td>
<td>5.03</td>
<td>4.88</td>
</tr>
<tr>
<td>15. Med. Rlz.</td>
<td>4.71</td>
<td>4.02***</td>
<td>4.00</td>
<td>3.44***</td>
</tr>
<tr>
<td>16. Lrg. Rlz.</td>
<td>3.89</td>
<td>3.11***</td>
<td>2.90</td>
<td>2.64</td>
</tr>
</tbody>
</table>

In each pair of compared structures (Tables 1 & 2, rows 1-12) three of the four parameters, the implicative intervals’ magnitude and both magnitude and direction of the realized ones, are held constant. In one member of the pair the implicative interval is ascending; in the other, it is descending. I also compared the mean ratings of ascending and descending realized intervals, regardless of the implicative intervals preceding them.
Separate comparisons were done for small, medium, and large realized intervals (rows 14-16 in Tables 1 & 2).

Paired t-tests determined whether differences between average scores for ascending and descending implications are significant. I computed the average scores of ascending and descending pairs for each subject, then computed for each subject the difference between these scores, and finally tested whether the average difference was significantly different from zero.

**Results**

Table 2 presents average scores for compared pairs in the four experiments. Scores in all experiments but Krumhansl and Schellenberg’s 3rd (Chinese) point at similar tendencies: they present highly significant preferences for ascending realizations, particularly when realizing intervals are medium or large (rows 13-16); correspondingly, continuity of direction is strongly preferred when the implicative interval ascends, and directional change—when implicative intervals descend (rows 1, 2). In particular, ascending Processes (P) and Intervalic Reversals (IR) are rated significantly higher than descending ones (rows 7, 11), while Registral Returns (IP/ID; row 8), in which the realizing pitch returns to the vicinity of the initial pitch, are preferred when concave (descending-ascending).

Preferences concerning change of interval size (rows 3-6) are weaker and not as consistent as those concerning change of direction. Notable is a tendency to rate increased magnitude higher when large implicative intervals descend (row 4b; see also VR, rows 12-12b). Correspondingly, in some experiments magnitude decrease is rated higher when large implicative intervals ascend (row 5b; see also Intervalic Reversal, row 11).

Tendencies in Krumhansl and Schellenberg’s “Chinese” experiment (Table 2, two leftmost columns) are very different from those in all other experiments. Most notably, this experiment presents preferences for descending, rather than ascending realizations, particularly when realizations’ magnitude is small (rows 13, 14). Correspondingly, descending Processes and Registral Processes (rows 7, 9) and ascending Registral Reversals (row 12) are preferred.

**Discussion**

**A. Differences Among Studies.**

The contrasts between results in Schellenberg and Krumhansl’s 3rd experiment and other findings could perhaps be related to the Non-Western origin of the stimuli in this experiment, and thus suggest that directional tendencies represent learned, culture-specific expectancies, rather than innate or universal ones. However, the peculiarities of the Chinese language, rather than any general East-West dichotomies, may have affected directional preferences here. Chinese is a tone language, where changes in pitch height affect a word’s meaning. Thus, specific constraints of language on pitch direction may have outweighed any musical constraint in this sample.
Though the generality of results reported here has yet to be determined, a forthcoming reanalysis by the present author of two additional melodic expectancy studies, Thompson et al. (in press), and Carlsen (1981), concurs with findings in Cuddy and Lunney and in Schellenberg and Krumhansl’s first two experiments. Hence, I rely on these findings in the discussion below.

B. Reevaluating the I-R model

Both Narmour’s original I-R principles (1990) and revised principles suggested by Schellenberg (1996, in press) seem to be affected by the implicative interval’s direction. Of Narmour’s principles, Registral Direction is strongly affected: for both small and large intervals, directional continuity is significantly associated with ascending implicative intervals, and directional change—with descending ones. Furthermore, while Narmour’s principle maintains that both ascending and descending small intervals imply directional continuity, our results show that when small implicative intervals descend, subjects prefer direction change consistently (see Table 2, rows 1a, 2a).

Schellenberg’s revised I-R model presents three principles, modifying Narmour’s original rules. The first principle, Proximity, replacing Narmour’s Intervallic Difference, maintains that smaller intervals are more strongly implied than larger ones, regardless of the implicative interval; the second principle implies Registral Return; the third revises Narmour’s Registral Direction principle by delimiting its application to large implicative intervals. All three principles seem to be affected by pitch direction. Registral Return tendencies are significantly stronger when implicative intervals descend (Table 2, row 8). Correspondingly, the revised Registral Direction principle seems to have a stronger effect when implicative intervals descend (rows 1b, 2b); and the Proximity implication is affected as well, since medium and large ascending realizations, denying proximity, are rated significantly higher than medium and large descending ones (rows 15, 16).

These differences between ascent and descent suggest that the basic registral symmetry, characterizing both Narmour’s original I-R model and its revised formulations, should be replaced by a direction-sensitive model, in which implications of ascending and descending intervals would be quantified and categorized differently. In such model, the nature and strength of a melodic implication would be affected not only by the implicative interval’s magnitude, but also by its direction.

C. Factors Affecting Directional Implication

Considered in view of other studies of registral direction, the present analyses suggest several possible factors influencing directional preferences.

Interval Magnitude correlates with pitch direction. Large and medium ascents are more strongly expected than large and medium descents, since ascending intervals larger
than a major third are more frequent in musical repertories than descending ones of a comparable size (see Vos & Troost 1989).

**Effect of rhythmic accents.** Test (realized) tones in the experiments were usually metrically and agogically accentuated, relative to the note preceding them. Studies of melodic accent (Thomassen 1982, Huron & Royal 1996, Eitan 1997) suggest that melodic rise may, in some circumstances, be correlated with metric and agogic accent. Thus, the rhythmic design of the stimuli may have biased subjects toward pitch rise.

**The pitch rivalry hypothesis.** As Schellenberg’s analyses (1996, in press) suggest, an implicative interval generates two rival implications of proximity, applying to the two tones comprising it. I propose that the relative salience of the “rival” notes affects the proximity conflict between them. Since higher pitches are more salient than lower ones, proximity to the higher tone of the implicative interval is enhanced. Thus, if the implicative interval descends, expectancies for direction change and for proximity to its first tone (Registral Return) are enhanced.

Findings reported in this paper suggest that pitch direction strongly affects melodic expectancy, and should be considered as a central factor in any model of melodic implication. A conclusive evaluation of the role of pitch direction in melodic implication should, however, await empirical examinations of the factors affecting directional expectancies; hopefully, this paper would encourage such endeavors.

**References:**


Music experiences may be strongly affected by implicit memory. Crowder (1993, p. 134) argues that although some melodies are remembered explicitly, "...the more ecologically general kinds of melody memory may occur implicitly ... the re-entry of a fugue theme, the occurrence of a leitmotif, or the developmental section of a remembered sonata-movement subject may all be examples of implicit music memory." In three experiments, we examined the influence of exposure to novel melodic fragments on judgements of melodic continuation. To anticipate our results, exposure to melodic fragments influenced judgements, and the influence went beyond what could be explained by conscious memory of the fragments. We argue that implicit memory for recently heard melodic patterns influences melodic expectancies.

Experiment 1

Experiment 1 was conducted to assess whether exposure to melodic sequences influences subsequent judgements of melodic continuation. Listeners were presented a series of short melodic sequences to both ears, and were required to record the number of notes in each. Because listeners attended to the number of notes, rather than to the patterns of pitch and duration, explicit memory for sequences should be minimal. After the exposure phase, we assessed the influence of exposure on judgements of melodic continuation.

Method

Participants. Twenty adults from a University student population (mean age = 23) volunteered to participate. All participants reported normal hearing. The number of years of music training of participants ranged from zero to greater than ten, with an average of six years experience playing a musical instrument. Participants were naive with respect to the purpose of the experiment, and were not familiar with the sequences.

Music sequences. Two sets (A and B) of 30 non-tonal melodic sequences, varying from 4-9 notes, were composed for the experiment. The duration of each note in all sequences was 0.30 s. Each sequence in Set A is identical to a sequence in Set B except for the final note. The number of semitones between the last two notes of corresponding sequences in Sets A and B is identical, but the direction of pitch movement is opposite. For example, if the final two notes of a sequence in Set A were D4 - E4, then the final two notes of the corresponding sequence in Set B would be D4 - C4. The number of sequences with descending and ascending endings is the same in the two sets.

Apparatus. MIDI files for the 60 sequences were created using Encore software (version 3.0.5). The tempo of all sequences was 100 quarter-tone beats per minute, or 0.30 s per (eighth) note. Presentation of sequences was controlled by a Power-Macintosh computer, using customized software (Experiment Creator). The computer was connected to a Roland Sound Canvas with equal temperament tuning. All notes were the sampled piano sounds (piano 1) of the Roland Sound Canvas. The loudness for all notes was set to the MIDI key-press velocity of 64. Participants heard the sequences through Sennheiser HD-480 Classic headphones, and were allowed to adjust the loudness to a comfortable listening level (approximately 70 dB SPL).

Procedure. Participants were tested individually. There were two phases: an exposure phase and a test phase. Participants took part in the exposure phase, followed by a pause of 10
minutes, and then the test phase. Participants were not informed of the test phase until after the exposure phase was completed. In the exposure phase, the 30 sequences in Set A were presented a total of three times. Sequences were presented in three blocks, and the order of presentation was scrambled within blocks. The 90 sequences were each separated by a 0.90 s pause. Participants were required to count the number of notes in every sequence, and to record each number on a form. In the test phase, all 60 sequences (Set A and Set B) were presented in an order that was randomly and independently determined for each participant. Participants were required to rate the degree to which the final note of each sequence was expected. That is, they were required to rate whether the final note was a predictable continuation of the preceding notes, in a musical sense. Ratings were entered directly into the computer using a scale from 1 to 7, where a rating near 1 indicated that the final note was unexpected, and a rating near 7 indicated that the final note was highly expected.

Results

As predicted, ratings of melodic continuation were higher for the set of sequences presented in the exposure phase (mean rating = 4.46) than for the set of sequences not presented in the exposure phase (mean rating = 4.11), F(1, 19) = 6.50, p < .02. The finding suggests that exposure to non-tonal melodic sequences influenced melodic expectancy. The finding illustrates a highly specific effect, because matching sequences in the two sets were identical except for the final note.

The analysis also revealed a significant effect of Sequence, F(29, 551) = 3.26, p < .01, and a significant interaction between Sequence and Set, F(29, 551) = 2.68, p < .01. The effect of Sequence illustrates that regardless of exposure, ratings of melodic continuation were higher for some sequences than others. The interaction between Sequence and Set indicates that mean differences between sequences depended on the Set of sequences considered. Multiple regression was adopted to explore the latter effects. Mean ratings of melodic continuation for each of the 60 sequences were modeled as a function of numerous predictors. Stepwise regression resulted in a model involving three significant predictors: Exposure (b = 0.35, t = 2.35, p < .03), Pitch proximity (b = -0.077, t = -2.48, p < .02), and Sequence length (b = -0.17, t = -3.23, p < .01).

Exposure predicted higher ratings for sequences presented in the exposure phase (coded as 1) than sequences not presented in the exposure phase (coded as 0). The predictive power of Exposure corroborates the results of the ANOVA, and reveals that exposure to sequences influenced judgements of melodic continuation beyond what could be attributed to Pitch proximity or Sequence length. Pitch proximity coded the number of semitones between the last two notes of each sequence. The negative weight associated with this predictor indicates that higher ratings of melodic continuation were assigned to notes that were most proximal in pitch to the immediately preceding note. This finding is consistent with studies of expectancy (e.g., Cuddy & Lunney, 1995; Krumhansl, 1995; Schellenberg, 1996; Thompson, 1996; Thompson, Cuddy & Plaus, 1997). Sequence length coded the number of notes in each sequence. The negative weight associated with this predictor indicates that participants assigned lower ratings of expectancy as sequence length increased. Possibly, as the length of a sequence increased, listeners were less able to assimilate that sequence into existing schemata, because sequences were composed in a non-tonal, unconventional (and hence unpredictable) style. The multiple correlation for the regression model, 0.52, is highly significant, F(4, 55) = 6.26, p < .001.

Experiment 2

The results of Experiment 1 suggest that repeated exposure to short melodic sequences influenced judgements of melodic continuation. In Experiment 2, we evaluated an alternative possibility. Although matching sequences in the two sets were identical except for the final note, it is possible that continuation notes in Set A (exposed set) would have been assigned higher ratings overall than those in Set B even without exposure. To rule out this possibility, participants in Experiment 2 provided ratings of continuation for all 60 sequences without prior exposure to either set.
Method

Participants: Fifteen adults fitting the description of participants in Experiment 1, and with a similar level of music experience, volunteered for the experiment. Participants were naive with respect to the purpose of the experiment, and none had taken part in Experiment 1.

Sequences. The sequences were identical to those used in Experiment 1.

Apparatus: The apparatus was identical to that used in Experiment 1.

Procedure: Participants took part in the test phase only, as described for Experiment 1.

Results

As expected, ratings of melodic continuation were almost identical for sequences in Set A (mean = 4.49) and Set B (mean = 4.50), F(1, 14) = 0.007, ns. This result suggests that the difference in ratings observed in Experiment 1 resulted from exposure to sequences in Set A.

As in Experiment 1, there was a significant effect of Sequence, F(29, 406) = 2.62, p < .01, and a significant interaction between Sequence and Set, F(29, 406) = 2.05, p < .01. Multiple regression was conducted to explore the latter effects. Mean ratings of melodic continuation for each of the 60 sequences were modeled as a function of the same predictors examined in Experiment 1. Stepwise regression again revealed significant predictive power by Pitch proximity (b = -0.10, t = -3.53, p < .01), and marginally significant predictive power by Sequence length (b = -0.09, t = -1.78, p = .08). The multiple correlation for the regression model, 0.43, is significant, F(2, 57) = 7.34, p < .01. Thus, except for the influence of exposure, ratings of melodic continuation were influenced by similar factors in Experiments 1 and 2, and there is a significant correlation between mean ratings of the 60 sequences in Experiment 1 and 2, r = 0.60, t(58) = 5.64, p < .01.

Experiment 3

The results of Experiment 1 and 2 suggest that exposure to non-tonal melodic fragments influenced judgements of melodic continuation. Experiment 3 was conducted to assess explicit memory for sequences in the exposure phase. Our aim was to evaluate whether the influence of exposure on ratings of melodic continuation could be attributed to explicit recognition of melodic sequences, or whether the effect was an instance of implicit memory.

Method

Participants: Twenty adults fitting the description of the participants in Experiments 1 and 2, and with a similar level of music experience, volunteered for the experiment. Participants were naive with respect to the purpose of the experiment, and none had participated in Experiment 1 or 2.

Sequences. The sequences were identical to those used in Experiment 1.

Apparatus: The apparatus was identical to that used in Experiment 1.

Procedure: Participants took part in the exposure phase, followed by a pause of 10 minutes, and then the test phase. The exposure phase was identical to that described for Experiment 1. In the test phase, all 60 sequences (Set A and Set B) were presented in a random order. Participants indicated whether they had heard each sequence during the counting task (the exposure phase). Judgments were entered into the computer by selecting "YES" or "NO" on the computer screen.
Results

Participants reported recognizing an average of 13.10 of the 30 sequences presented in the exposure phase (Set A), and 10.00 of the 30 novel sequences (Set B). This difference was reliable and indicates that participants retained a small degree of explicit memory for sequences in the exposure phase, $t(19) = 4.15, p < .02$. Unlike ratings of melodic continuation, there was no relationship between recognition accuracy and sequence length. The probability of making a correct response for each of the 60 sequences (i.e., responding yes for exposed sequences, or no for unexposed sequences) was uncorrelated with sequence length ($r = 0.09, df = 58, ns$).

Multiple regression was adopted to examine whether explicit memory for sequences may have influenced ratings of melodic continuation observed in Experiment 1. Mean ratings of the 60 sequences in Experiment 1 were regressed on three predictors: Mean ratings of the 60 sequences when there was no exposure phase (Experiment 2); the probability of making a recognition (i.e., "yes") response when there was an exposure phase (Experiment 3); and Exposure to the sequences. The latter predictor was dummy coded: a 1 was assigned to sequences that were presented in the exposure phase and a 0 was assigned to sequences that were not presented in the exposure phase. If the exposure effect observed in Experiment 1 were based merely on explicit memory for sequences, then the actual exposure of sequences should add no predictive power beyond that provided by recognition responses. Conversely, if judgements of melodic continuation reflect implicit memory for previously heard sequences, then exposure to sequences should have predictive power beyond that provided by recognition responses.

Stepwise regression resulted in a model with two significant predictors: Mean ratings from the control group ($b = 0.66, t(57) = 6.00, p < .01$), and Exposure to sequences ($b = 0.35, t(57) = 2.68, p < .01$). The multiple correlation, 0.65, is highly significant $F(2, 57) = 21.20, p < .01$. Recognition responses provided no predictive power beyond that provided by these predictors (partial correlation = 0.04, ns). Moreover, the correlation between Recognition responses and mean ratings of continuation (Experiment 1), 0.08, is not significant, $t(58) < 1.0, ns$. These results suggest that exposure to sequences influenced ratings of melodic continuation beyond what can be explained by explicit memory for sequences. That is, judgements of melodic continuation reflected implicit memory for previously heard melodic sequences.

Discussion

Experiment 1 and 2 indicate that exposure to non-tonal melodic fragments has a reliable influence on judgements of melodic continuation. Listeners assigned higher ratings of melodic continuation to sequences presented in the counting task than to novel sequences. The findings extend Wilson's (1979) study of the effects of recent exposure to melodic stimuli. However, whereas Wilson reported an effect of exposure on preference judgements, our study illustrates an effect of exposure on expectancy. This influence of exposure on expectancy may be instantiated often in music experiences. In particular, composers may exploit such effects in their presentation of theme and variation. As expectancies for a thematic pattern are established and strengthened through repeated presentation of a theme, composers may introduce variations on that theme, knowing that listeners will assimilate both theme and variation into an established schema.

Examination of recognition data suggested that judgements of melodic continuation were not based on explicit memory for sequences, even though they were influenced by exposure. The effect of exposure may therefore be interpreted as a manifestation of implicit memory. That is, melodic expectancies may be partially based on implicit memory for recently heard melodic events. These findings contribute to existing data on implicit memory, and provide one of the few demonstrations of implicit memory for melodic stimuli (for a review of research, see Schacter, 1987).

The findings also are relevant to research on melodic expectancy (e.g., CarlSEN, 1981; Jones, 1981, 1982, 1990, Unyk & CarlSEN, 1987). Narmour (1990, 1992) identified three classes of
influence on melodic expectancy: innate or universal influences; long-term knowledge of music; and memory for recently heard patterns in the current melody, or intraopus style. Studies have addressed the first two classes of influence (e.g., Cuddy & Lunney, 1995; Krumhansl, 1995; Schellenberg, 1995; Thompson, Cuddy & Plaus, 1997). The current results illustrate the third class of influence on melodic expectancy. Exposure to melodic sequences led to an expectation that similar patterns of continuation would occur. Our results extend Narmour's discussion of intraopus style, however, by showing that influences by recently heard passages may be implicit.

Finally, the current investigation provides clues as to how listeners might develop an appreciation for non-tonal or unconventional musical styles. Our results indicate that melodic expectancies are affected even by limited exposure to novel melodies composed in an unfamiliar style. With repeated exposure to melodies composed in an unfamiliar but consistent style, listeners may develop stable expectancies for similar melodic patterns. As such expectancies become widespread among listeners, composers may then create patterns of denials and fulfillments of melodic implications, shaping listeners' aesthetic responses as a melody unfolds.

References


Acknowledgements: This research was funded by a research grant from the Natural Sciences and Engineering Research Council of Canada, awarded to the first author.
Tapping in time with musical and mechanical sequences
Carolyn Drake*, Amandine Penel*, Emmanuel Bigand** and Luca Stefan*
*Laboratoire de Psychologie Expérimentale, CNRS URA 316, 28 rue Serpente, 75006 Paris, France, e-mail: drake@idf.ext.jussieu.fr
**LEAD, CNRS, Université de Bourgogne, Dijon, France

Tapping one's foot in time with music is the most common and obvious physical manifestation of music perception, if one excludes dancing. This task is relatively easy for most people, it depends on the music, and is sometimes difficult to inhibit. As such, it shows that the extraction of temporal regularities to identify the underlying beat is a process fundamental to music understanding. As cognitive psychologists, we are interested in the processes that allow listeners to accomplish this task. The Dynamic Attending Theory (Jones & Boltz, 1989) provides an on-line explanation. Synchronisation is possible because listeners can attune their internal rhythms to those in the environment. It is proposed that listeners spontaneously focus on an intermediate rate (hierarchical metric level in the case of music) called the reference period. One reflection of this reference period would be the rate at which one taps one's foot in time with the music. Once attuned to a particular level in the metric hierarchy, listeners can focus on other levels, through a process of focal attending, either to multiplications of the beat (future-oriented attending) or to subdivisions of the beat (analytic attending). The first aim of this paper is to provide experimental support for this theory by demonstrating that each of these processes is involved in the task of synchronizing with musical sequences. The comparison between the abilities of musicians and non musicians participants throws light on the way in which skill influences the functioning of, and the interaction between, these processes.

The second question we address is the role of temporal regularity on the ability to synchronise with sequences, on the referent period and on focal attending. Beat extraction has been studied by tapping tasks in expressionless sequences (Parncutt, 1994; Vos, 1994), but how is beat extracted in natural expressive sequences? Most beat and meter induction models (Longuet-Higgins & Lee, 1982; Povel & Essens, 1985; Desain, 1992; Parncutt, 1994) predict that the more regular a musical sequence is, the easier it is to extract the underlying beat (which makes sense). They therefore predict that it should be easier to synchronise with mechanical performances synthesised by computer than with real performances which contain many expressive temporal microvariations, as well as variations in intensity, articulation and pedal. However, a more cognitive approach, considering that performers use these microvariations to highlight aspects of the structure and to transmit a particular interpretation to the listeners (Drake & Palmer, 1993; Palmer, 1989; Sloboda, 1985) would be less categorical in its predictions: these additional cues would be used by the listeners to elaborate a more complex mental representation of the sequence, which may lead to facilitated synchronisation with real performances compared to mechanical performances.

In this study, we examine the way in which musicians and non musicians synchronise their taps with different versions of musical sequences: mechanical, expressionless versions (synthesised by computer), accented versions (mechanical versions containing accents on the first beat of each measure), and natural versions containing many types of expressive microvariations. The comparison between performances for the three versions will enable us to distinguish between the predictions of beat extraction models and those of a more cognitive approach.

METHOD

Participants. 36 participants took part in the experiment, with a mean age of 24 years (ranging from 20 to 28 years). Among them, 18 were non musicians and 18 musicians. Non musicians had never received any formal music training or played any instrument, whereas
musicians had an average of 7 years of lessons (minimum = 5), 14 years of practice and 8 hours of practice per week.

Materials. Six pieces for piano were used, covering a wide range of different styles encountered in Western tonal music: the minuet I from Partita No.1 by J.S.Bach, a movement of the Sonata in H (K.261-L.148) by D.Scarlatti, the first movement of a Leichte Sonata (op.49, No.2) by L. van Beethoven, a Mazurka (op.7, No. 1) by F.Chopin, a Rumanische Volkstanz (Der Tanz mit dem Stabe) by B.Bartok and the "Menuet sur le nom de Haydn" by M.Ravel.

Three versions of each of these pieces were used. The natural version was a real performance by a professional pianist. It thus contained many performance microvariations such as timing, intensity, articulation and pedal variations. The mechanical version was synthesised by computer. All expressive variations were absent, i.e. the sounded result corresponded exactly to the written score. The tempi corresponded to the mean tempi used by the pianist and, for each piece, a mean intensity was defined by computing an arithmetic mean over the midi velocities of the notes in the natural version. Articulation was defined for each note by its duration in the score (hence, everything was played legato). No pedal was used. The accented version was a synthesised mechanical version in which metric accents were added by increasing by 10dB the intensity of notes situated on the first beat of each measure.

Apparatus. The recordings of the natural versions where carried out on a Yamaha Disklavier II monitored by a computer using the sequencer Master Tracks Pro4. In the same silent room, these natural versions and the mechanical and accented ones were played back on the Disklavier using Master Tracks Pro4, line out. To tap in time with the music, participants used a Roland Pad PD7 connected to a sound synthesizer Roland TD5 and to the computer for the recording of the taps in Master Tracks Pro4, line in.

Procedure. Each participant heard all six pieces, each in one of the three versions. The two pieces heard in each version varied systematically between participants. The order of the versions and of the pieces was counterbalanced. For each piece, the participants first listened to the music to familiarise themselves with it. They could listen to it several times or interrupt the listening phase before the end of the piece. Then, during the production phase, they were asked to tap on the Pad with their dominant hand, regularly and in synchrony with music. In a first phase, they were asked to tap at their spontaneous speed, i.e. at the speed that seemed most appropriate to the sequence. For half of the participants, in a second phase, they were asked to tap, if they could, faster than their spontaneous speed, but still staying in time with the sequence, and then if they could faster still, and so on. In a third phase, they were asked to tap, if they could, slower than their spontaneous speed, but still staying in time with the sequence, and then if they could slower still, and so on. The other half of the participants did the opposite order (slower and then faster). The experimental session lasted about half an hour.

Dependent measure. A participant was considered to have successfully synchronised a measure if all the taps occurred at regularly spaced intervals corresponding to a particular metrical level (+/- 10%).

RESULTS

1. Ability to synchronise

The ease of initial synchronisation was investigated by examining the number of measures out of the first ten measures (measures 2 to 11 as there was no indication to signal the beginning of the pieces) for which the participants successfully synchronised. Overall synchronisation was good (musicians = 8.9 and non musicians = 7.5). Figure 1 presents the number of successfully synchronised measures as a function of the version. Initial synchronisation was easier for musicians than for non musicians. It was easier with mechanical or accented versions than with natural versions.
An analysis of variance (ANOVA) on the number of successfully synchronised measures by group (musicians and non musicians) and version (mechanical, accented and natural) confirmed significant effects of group \((F(1, 210) = 20.5; p<.0001)\) and of version \((F(2, 210) = 9.9; p<.0001)\), but no interaction between the two. Initial synchronisation was significantly more difficult with natural than with mechanical and accented versions \((F(1, 210) = 18.7; p<.0001)\), which did not differ significantly.

The participants may have had difficulty establishing the synchronisation, but how well could they synchronise once they had got going? To address this question, we examined the number of measures for which the participants successfully synchronised out of the ten successive measures in the excerpt for which they most successfully synchronised. Overall, synchronisation was of course better than before (musicians = 9.8 and non musicians = 8.8), with a ceiling effect for musicians. Otherwise, the pattern of results was very similar.

Thus, initial synchronisation was easier and synchronisation better for musicians than for non musicians. For musicians as well as for non musicians, the temporal regularity provided by mechanical and accented versions facilitated both measures, compared with the microvariations present in natural versions. The presence of metric accents did not significantly improve performances on these tasks compared with the mechanical versions.

2. Spontaneous synchronisation tempo

The rate at which each participant synchronised his taps with the music was measured in the following way: 1) we established the ten successive measures for which the participant tapped most regularly, 2) each tap was coded in relation to its corresponding position in the measure of the music in order to identify the metric level at which the participant tapped (eighth-, quarter-, half-notes etc.), 3) the mean tapping rate (in milliseconds) was then calculated in relation to the performed tempo of each piece. Figure 2 presents the spontaneous synchronisation tempo as a function of the version. On average, musicians tapped slower than non musicians. Also, musicians and non musicians tapped slower with accented or natural than with mechanical versions.

![Figure 1: Number of successfully synchronised measures as a function of the version.](image)

![Figure 2: Spontaneous synchronisation tempo as a function of the version.](image)
An ANOVA on the spontaneous synchronisation tempo by group and version confirmed significant effects of group \((F(1, 195) = 6.1; p<.02)\) and of version \((F(2, 195) = 4.3; p<.02)\). Despite the lack of significant interaction between group and version \((p=.17)\) significantly different profiles were observed for the two groups of participants. For musicians, the spontaneous synchronisation tempo was significantly slower for the natural than for the accented and mechanical versions which did not differ significantly. For non musicians, the spontaneous synchronisation tempo was significantly faster for mechanical than for accented and natural versions which did not differ significantly. These results can be summarised in the following way:

- **Musicians:** Nat>Acc, Mec
- **Non musicians:** Acc, Nat>Mec

Thus, for both groups, temporal regularity led to a fast spontaneous synchronisation tempo. The slowest tempo was obtained for non musicians with accented versions, and for musicians with natural versions.

For mechanical and natural versions, musicians tapped significantly slower than non musicians \((F(1, 195) = 3.6; p<.06; F(1, 195) = 5.9; p<.02\) respectively). However, there was no significant difference between musicians and non musicians for accented versions. This suggests that, for accented versions, the metric structure of the pieces was perceived the same way by musicians and non musicians. This interaction between group and version was further investigated by examining the metric level at which participants tapped in relation to the meter of each piece (there were two pieces in 2/4, one piece in 4/4 and three pieces in 3/4). Five levels were considered: below the quarter-note level, the quarter-note level, between the quarter-note level and the measure level, the measure level, and above the measure level. Figure 3 presents histograms of tapping rates for the three versions.

In general, tapping levels follow a bimodal distribution: most of the participants tapped either at the quarter-note level or at the measure level. In agreement with the previous analysis, musicians tended to tap at a higher metric level than non musicians (measure and quarter-note respectively). This general pattern varied with the versions. For the mechanical versions, non musicians tended to tap at the quarter-note level, whereas more musicians tapped at the measure level. Thus, musicians spontaneously focused higher into the musical structure than non musicians who remained close to the musical surface. For the accented versions, both groups tapped more at the measure level, indicating the effectiveness of the added intensity accents to highlight the measure level. Non musicians were able to use this cue to synchronise with a higher hierarchical level. For the natural versions, a wider range of tapping was observed non musicians, suggesting a greater incertitude about the appropriate hierarchical level.
3. Number of synchronisation levels

After having synchronised with the music at their spontaneous synchronisation tempo, participants were asked to synchronise faster and faster still, and slower and slower still. We examined the number of levels at which they were able to synchronise. Figure 4 presents the number of levels as a function of the version. On average, musicians were able to synchronise with more levels than non musicians. Also, musicians and non musicians were able to synchronise with more levels with mechanical or accented than with natural versions.

An ANOVA on the number of levels by group and version confirmed significant effects of group \( (F(1, 210) = 53.4; p<.0001) \) and of version \( (F(2, 210) = 18.5; p<.0001) \), but no interaction between the two. Musicians synchronised with 4 levels on average (ranging from 1 to 7) and non musicians with 3 (ranging from 0 to 6). The number of levels was significantly lower with natural than with mechanical and accented versions \( (F(1, 210) = 37.0; p<.0001) \) which did not differ significantly.

Thus, musicians had access to a larger range of synchronisation than non musicians. Temporal regularity (with and without metrical accents) led to a larger range of possible synchronisations.

**DISCUSSION**

This experiment confirms what we already know - musicians are faster and better able to synchronize their taps with musical sequences. Of more interest is the demonstration that musicians and non musicians do not apprehend musical sequences in the same way: musicians tap spontaneously at a higher hierarchical level than non musicians, and have access to a larger range of hierarchical levels. These results can be interpreted in the light of the Dynamic Attending Theory: (1) Musicians are faster and better able to synchronise their taps with sounded sequences than non musicians because they have enhanced attunement skills (the ability to synchronise their internal rhythms with those in sounded sequences). This is probably due to the ability to focus preferentially on relevant information. (2) Musicians spontaneously tap at higher hierarchical levels than non musicians, suggesting that they have a slower reference period. (3) Musicians are able to tap at more hierarchical levels than non musicians, suggesting greater abilities to focally attend. They therefore have a more elaborate mental representation, further from the musical surface, more hierarchically organised and more flexible.

A second aim was to examine the role of expressive microstructure on listeners' ability to follow the musical sequences. Does the presence of expressive microvariations enhance (as predicted by a cognitive approach) or inhibit (as predicted by models of beat induction) these abilities? Two abilities (the ability to synchronise and the ability to pass from one hierarchical level to another) are inhibited by the presence of microvariations in the natural versions for both the musicians and non musicians: a strict temporal regularity enhances spontaneous synchronisation and synchronisations at other hierarchical levels. However, the pattern of
results observed for the spontaneous synchronisation tempo suggests that expressive microstructure has a considerable influence on the way listeners apprehend the sequences. Spontaneous synchronisation is at a higher hierarchical level when microvariations are present, with some participants synchronising for example at the beat level with mechanical versions and at the measure level with accented or natural versions. This can be interpreted in the light of a coding efficiency concept. Because of perception and memory limitations, in order to make sense out of the multitude of sounds arriving to their ear, listeners have to code information in some efficient way (for example, hierarchical organisations achieve a reduction of information coding load). Expressive musical sequences present microvariations which highlight some aspects of the musical structure (Palmer, 1989). They thus reduce information coding load by creating more coherent mental representations of the music, and result in the perception of larger units and in a slower spontaneous synchronisation tempo than with expressionless sequences. The effect of intensive microvariations, and more precisely of metric accents, was tested with the use of accented versions. The results show that metric accents improve coding efficiency for both musicians and non-musicians, but that other microvariations further improve coding only for musicians. These effects of the version on the spontaneous synchronisation tempo can also be explained by the fact that metric accents orient perception at the measure level, as do expressive microvariations present in the natural versions (which certainly include metric accents). The examination of the number of levels to which participants can synchronise shows that it is lower for natural versions than for mechanical and accented ones, which do not differ. Thus, metric accents do not restrict perception as much as do all the microvariations present in natural performances, as they allow more passage to other levels. This highlights the success of the natural performances to convey a particular interpretation of the music to the listeners.

REFERENCES


Impact of Melodic Structure on Emotional Expression

Erik Lindström
Department of Psychology, Uppsala University, Sweden

Introduction
Theories concerning music and emotion often refer to a close connection between emotion and dynamic aspects in music. The aesthetician Stephen Davies postulates that emotion characteristics presented in music depend on its dynamic character; "for example on patterns of tension and release" (Davies, 1994, p. 244). Music's ability to express emotions may be achieved using many different means. In an on-going research project in the Department of Psychology at Uppsala University we investigate the means that musicians use when they play the same piece of music with different emotional characters. These studies have demonstrated that practically all performance variables - such as tempo, timing, articulation, loudness, amplitude envelopes, timbre - are affected by the different intentions concerning the emotional expression (Gabrielsson & Lindström, 1995; Gabrielsson & Juslin, 1996). However, besides the fact that musicians may affect the emotional expression solely by their performance, it also seems appropriate to consider the musical piece in itself as possessor of certain expressive characteristics. That is, a piece of music with a certain structure has an inherent expression that the performer is to administer.

Empirical studies by Hevner (1936), Rigg (1964) and Wedin (1972) and theoretical work by Cooke (1959), Meyer (1956) and Narmour (1991) indicate the impact of various factors related to the musical piece on emotional expression. However, there seems to be no investigation, that, with factors in the performance under control, in more detail examines the effects on emotional expression of melodic/harmonic structure, melodic contours and rhythm.

"Tension" and "resolution" are often mentioned in theoretical work on tonal structure and hierarchy. Pitches are assumed to be hierarchically organized in a system with a central reference pitch. The dynamic vehicle within this structure is that unstable tones are pulled towards more stable tones (Krumhansl, 1990). Formations of tone sequences create implicit harmonic functions, that is, chords or harmonic progressions. Some progressions are clear and definite, whereas others create feelings of vagueness and ambiguity.
Melodic contours and rhythm patterns may as well contribute to certain expressive characteristics in music. However, they may also be decisive in the creation of tonal structure and hierarchy. Theoretical and empirical works indicate the importance of rhythm and temporal accentuations in harmonic functions (Boltz, 1993; Lerdahl, 1988). Other studies demonstrate that specific positions and orderings of tones determine the existence of certain harmonic functions as well (Deutsch & Feroe, 1981; Povel & Egmond, 1993; Butler 1988; Brown, 1989).

Central to Meyer’s notions of emotion and meaning in music (1956, also cited in Krumhansl, 1990) were the concept of tonal structure and the generation of tension and resolution, instability and stability, ambiguity and clarity. He also discussed the importance of melodic features and rhythm. However, these issues are still largely unexplored and the purpose of this investigation was therefore to take some of these aspects under scrutiny, that is, to study if and to what extent harmonic function, melodic contour and rhythm pattern have an impact on emotional expression in music.

Method
72 different versions of the tune Frère Jacques were created. A listening test was conducted to see whether these different versions had an impact on emotional expression.

Participants
19 psychology students, 13 females and 6 males, aged 20 - 31 years, volunteered to participate in the experiment and were paid 140 SEK each. Only two of them had played a musical instrument.

Musical materials
The music sequences in this experiment were all derived from the well-known tune Frère Jacques. Six melodies that looked like the original tune concerning melodic lines and contours but that differed in their tonal progression, were created (Figure 1). The first three sequences are in major mode and diatonic (Figure 1a-1c), whereas the last three in minor mode show various degrees of violations to diatonicism (Figure 1d-1f). The intention was to create a variation extending from simple and stable harmony to complex and unstable harmony, accomplished by violations to diatonicism as well as by insertions of unusual harmonic progressions The next step was to systematically combine these “prototypic“ melodies with variation in melodic direction (“original“ and “reversed“) and contour (“original“ and “complex“). Changes in melodic directions were accomplished by use of reversed tone sequences within measures relative the original tune, transforming a rising sequence to a descending sequence. Relative the original tune, complex contours were created by more directional changes. The intention was to measure the impact of melodic motion, as well as the impact of serial orderings in harmonic materials. Furthermore, these sequences were rhythmically manipulated with regard to the ratio between the two notes in pairs of quarter notes (see figure 1). The ratios used were 1:1 as in the original tune, 3:1, long-short; and 1:3, short-long. The intention was to assess the impact of overall rhythmic structure as well as effects of rhythmic stress markings in different harmonic parts. A total of 72 sequences were created.
Figure 1. Notation of six “prototypic” versions of “Frère Jacques”. Sequences a-c in major mode (left) and sequences d-f in minor mode (right).

Apparatus and recording conditions
The sequences were all prepared in the midi-sequencer programs Cakewalk Express 3.0 and Cubase 3.0 for Windows. The recording of the sequences was made via midi to a synthesizer (Korg wavestation) using a saxophone-like timbre to a DAT recorder (Pioneer D-05). All sequences were programmed in a tempo of 120 bpm (beat = quarter note), and all tones were given the same loudness and articulation (the onset-offset duration was 75% of the inter-onset interval). In the experiment the sequences were reproduced at a comfortable sound level using the DAT recorder, a Pioneer amplifier and AR 12 loudspeakers.

Procedure
The experiment was conducted in a big lecture hall. The participants were divided into two groups. Each group listened to all 72 musical sequences in randomized orders, distributed over two sessions with one hour break between the sessions. They were told that they would listen to different melodies that all were variations on a famous tune, and that their task was to rate each melody on six scales (see Design). First a pretest was conducted using 8 sequences representative for the great variety of occurring sequences. In the second session the pretest was limited to only two sequences. Each session lasted for about 45 minutes, including a short break for a few minutes in the middle.

Design
The design was a 6 x 3 x 2 x 2 factorial within subjects design, involving six levels of tonal progression, three levels of rhythm, two levels of melodic contour, and two levels of melodic direction. The dependent variables were three bipolar scales comprising seven steps and reflecting perceived structure -“stable-unstable”, “simple-complex”, “relaxed-tense”, and furthermore three scales referring to emotional expression - "happy-sad", "tender-angry", and "expressionless-expressive".
Results

The manipulation of melodic materials had a clear impact on judgments in every scale. The effects on judgments of structural aspects were similar for all six progressions (Figure 2, left). Progression no. 3 is judged as the most "stable", "simple" and "relaxed", whereas the following progressions (nos. 4-6) show successively more "instability", "complexity" and "tension". Progression no. 2 is judged midway between occurring poles in those scales, that is, with scale degrees around 4. This is also the case for progression no. 1 (the original progression of Frère Jacques) but with tendencies for "stable", "simple" and "relaxed", in scale degrees around 3.5. With regard to the emotion scales (Figure 2, right), the results for "tender-angry" look similar to those of the structural scales. However, the judgments of "happy-sad" and "unexpressive-expressive" were somewhat different: progressions nos. 1-3 in major were judged as "happy" and progressions nos. 4-6, in minor mode, as "sad". The result for "expressionless - expressive" indicate a tendency for "expressive" in each case. ANOVAS showed significant main effects (p<0.001) among the six levels of tonal progression in each scale.

![Figure 2](image)

Figure 2. Mean ratings across subjects for progressions 1-6; structural scales (left) and emotion scales (right).

The manipulation of rhythm indicates that use of 3:1 and 1:3 ratios in most cases led to increased "instability" and "complexity" (p<0.01) but not "tension" in comparison with the 1:1 ratios (pairs of quartenotes) as in the original tune. Significant main effects of rhythm (p<0.05) also appeared in all emotion scales. The 3:1 ratios were judged as more "sad" and "tender" than the ratios 1:1 and 1:3. The 3:1 and 1:3 ratios were judged as more expressive than the 1:1 ratio.

Melodic contour and direction showed the smallest effects on every scale. "Instability" and "complexity" was slightly increased by more contour shifts (p<0.1). Reversed melodic directions were judged as more "tender" relative original melodic directions (p<0.01). However, these two variables interact with each other and with other variables as well. Figure 3a shows the interaction between contour and rhythm for "happy-sad"(p<0.01). With the
original contour there was no dispersion among rhythmic proportions, but this was the case with complex contour. Within this condition the 3:1 ratio was judged as more sad, 1:1 ratio as more happy and 1:3 ratio in between. Figure 3b shows the interaction between melodic direction and rhythm (p<0.01) in the tender-angry scale. The ratios 3:1 and 1:3 were judged as more "tender" in reversed melodic directions than in original directions. For 1:1 ratio this effect did not occur.

![Figure 3. Mean ratings across subjects for contour and rhythm in "happy-sad" scale (left) and for melodic direction and rhythm in "tender-angry" scale (right).](image)

**Discussion**

The result support the notion that factors in melodic structure have expressive functions. There also seems to be some connection between judgments of emotional expression and of structure. The six tonal progressions seem to have the strongest impact on the expressions. However, the occurring interactions demonstrate an intricate interplay between, for instance, melodic features (contour/direction) and rhythm in emotion scales. These interactions indicates the importance of melodic motion in expressive music and these issues will be further investigated.

Theories in musical aesthetics and musicology are often hard to test empirically. The above results suggest that there is a certain connection between melodic structures and emotional expression. Research on musical structure and on emotional expression have often been conducted separately from each other. It seems that an integration would be fruitful.

**Acknowledgments**

This research was supported by The Bank of Sweden Tercentenary Foundation through a grant to Alf Gabrielsson.
References


TEMPO MODULATION TASK
Marek Franěk, Institute of Musicology, AS CR, Prague
Jiří Mates, Institute of Physiology, AS CR, Prague

Still very little is known on human skill to modulate tempo. A special example of this skill is a task to perform a sudden tempo change in a given tempo proportion. Are subjects able to perform exact tempo modulation in ratios such as 1 : 2, 1 : 3, or 1 : 4? Are there any preferred ratios when subjects are asked to make an arbitrary tempo modulation? In our previous experiment (Franěk and Mates, 1996) we allowed the subjects to carry out arbitrary tempo modulation to seek for potential existence of any preferred ratios of tempo change. The results showed the necessity to focus our attention towards an investigation of ability to perform doubling/halving tempo change. Do musicians perform double tempo change exactly or not? Large differences from 2:1 ratio in doubling or halving of tempo have been found in a study on tempo performance in jazz (Collier and Collier, 1994).

The present experiment was designed to study doubling/halving tempo change during motor tapping task. Wide range of tempi was employed in order to find a possible influence of tempo speed range. Since the results of the study by Collier and Collier (1994) indicate that precision of doubling/halving change of tempo is influenced also by musical context, we decided to study tempo change in simple pulse tapping sequence.

METHOD

Subjects

Sixteen subjects participated. They were musical amateurs from 19 to 37 years old.

Stimuli and procedure

Subjects were asked to perform doubling/halving change of speed of their finger tapping. The task consisted of two periods: first, the establishment of the base tempo, then the introduction of tempo modulation. In the first period subjects heard a sequence of stimulus tones presented in base tempo and were asked to synchronize their tapping with the sequence. Upon termination of the stimulus sequence the second period of the task was initiated, at which the subjects were asked to change the tempo and to continue in a newly established speed of tapping. After 50 taps, subjects were asked to stop.

Stimulus sequence consisted of 20 tones of 100 msec duration at a frequency of 440 Hz and it was presented in 7 different base tempi. Tempo was defined by a duration of temporal interval between successive tone onsets (= beat interval). Following base tempi were used: 400, 500, 600, 700, 800, 900, 1000 msec. For instance, a subject was asked to speed up base tempo 400 msec to modulated tempo 200 msec or to slow down this base tempo to modulated tempo 800 msec. Each subject completed one session consisting of two blocks. Speeding up was performed in the first block, slowing down in the second block. Stimulus sequences in particular base tempi were presented in a random order. Subjects were asked not to make any body movements in order to avoid interference with tempo change task. After completing the two blocks subjects were asked, whether he/she felt a difference in...
difficulty of the task between speeding up and slowing down.

**Apparatus**

The stimuli were generated and the subject tapping was recorded by a PC program. Stimulus tones were generated by the PC’s internal sound generator and presented via PC’s loudspeaker. Subjects tapped with their index finger on a special tapping device.

**RESULTS**

Motor responses were measured from the moment the stimulus sequence terminated. The temporal intervals between successive taps (intertap intervals) were evaluated and recorded. In the data analysis the first 10 taps are omitted. Modulated tempo was expressed as the average over all intertap intervals of particular tapping sequence. Table 1 shows mean values of modulated tempi over all subjects. Mean values of tempo modulation for particular base tempi and expected values corresponding exact ratio 2:1 ratio are plotted on Figure 1.

<table>
<thead>
<tr>
<th>Base tempo</th>
<th>400</th>
<th>500</th>
<th>600</th>
<th>700</th>
<th>800</th>
<th>900</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed up</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>204</td>
<td>244</td>
<td>284</td>
<td>331</td>
<td>365</td>
<td>409</td>
<td>454</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>14.4</td>
<td>27.2</td>
<td>27.5</td>
<td>50.3</td>
<td>32.2</td>
<td>28.5</td>
<td>45.9</td>
</tr>
<tr>
<td>Slow down</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>839</td>
<td>1024</td>
<td>1189</td>
<td>1342</td>
<td>1474</td>
<td>1694</td>
<td>1883</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>96.9</td>
<td>64.1</td>
<td>77.3</td>
<td>122.6</td>
<td>70.2</td>
<td>58.7</td>
<td>65.6</td>
</tr>
</tbody>
</table>

Table 1: Mean intertap intervals in msec (Mean) of modulated tempi and standard deviations (Std. Dev.) for particular base tempi during conditions speed up and slow down. The data were averaged over all subjects.

It turns out that only in fast tempi the results of tempo modulation were close to mathematically exact ratio 2:1. However, in the other cases the values of tempo modulation deviated from expected ones. The statistical significance of differences between values was tested with T-test for dependent samples. In the following cases the differences were significant: Speed up/600 (t = -2.19894, p < .043984), Speed up/800 (t = -4.27976, p < .000658), Speed up/900 (t = -5.70165, p < .000042), Speed up/1000 (t = -4.13052, p < .000890), Slow down/800 (t = -2.94492, p < .010036), Slow down/900 (t = -2.65419, p < .018040), Slow down/1000 (t = -2.82192, p < .012877).
In further analysis relative deviations of modulated tempo from expected tempo were calculated as a ratio between the values of modulated tempo and corresponding expected tempo (see Fig.2). Two-way repeated measures analysis of variance was computed to find an effect of base tempo and direction of tempo change on value of tempo modulation. Base tempo and direction of tempo change were chosen as independent variables, the ratio between modulated and expected tempo as dependent variable. Results showed that the value of base tempo had significant effect ($F = 7.6963$, $p < .000001$), while the direction of tempo change did not. There was no significant interaction between effects of base tempo and direction of tempo change. In fast and intermediate tempi the clear effect of the speed of base tempo on the deviation from 2:1 ratio was found - the slower the tempo was, the greater was the deviation. In slow tempi deviation from the correct ratio did not further increase.

---

**Figure 1:** Mean values of tempo modulation. Points marks mean intertaps intervals of modulated tempi, dashed lines indicate corresponding values for exact 2:1 ratio.

**Figure 2:** Values of modulated tempi relative to expected tempi.
Most of the subjects indicated the task speed up as simpler than the slowing down task. However, analysis of the results did not support this subjective feeling.

DISCUSSION

The experiment had two main issues: (1) doubling/halving tempo change was not performed exactly; (2) base tempo influences the value of deviation from the 2:1 ratio. The first issue is in line with a large body of research results obtained in analysis of timing of music performance. Temporal structure of a piece of music is not performed in its mathematically exact form. On the contrary, in performed music many various "systematic variations" from an exact norm appear (see for instance Bengtsson, Gabrielsson, 1980). The present experiment shows systematic variation during doubling/halving tempo change. The second issue concerns psychophysics of time. For instance, it is known that the threshold for discrimination of tempo change is dependent on the speed of tempo (Schulze (1989, Drake et al., 1992). Some studies showed that the link between speed of tempo and the discrimination threshold is complicated (Ellis, 1991, Vos et al., 1996). Our results showed the interesting bias - in the majority of sequences in the task speed up the subjects tended to speed up a little bit more than twice, while in the task slow down they tended to slow down a little bit less. We can take into account an existence of an "ideal" range of tempi in both rhythmical and perceptual tasks (see for instance Fraisse, 1978). This range is between 400 to 800 msec (duration of beat interval). This might be a reason for the general tendency to tap modulated tempo faster then it might be. However, satisfactory explanation of observed phenomena requires further systematic research.

REFERENCES

Rhythm and time
- the influence of time per se on
the organisation of temporal patterns

Guy Madison
Department of Psychology
Uppsala University
Sweden

One of the more fascinating aspects of rhythm is that it provides a vehicle for the study of human time. Rhythm is formed and constrained by properties of the experiencing organism, and while various effects of the interval between the onsets of perceptual events have been reported (e.g. Fraisse, 1982; Friberg & Sundberg, 1995), there are very few studies of pattern duration (Madison, 1997). The fundamental problem is how perceptual events, scattered in time, are integrated by the mind to meaningful structure. In fact, the notion of a time-window comprising what is psychologically present is almost as old as psychology itself. It is typically addressed in papers with a more theoretical approach (e.g. Block, 1990; Michon, 1978; Pöppel, 1988; Pöppel, 1978), also with direct reference to music and poetry (Turner & Pöppel, 1988). Most empirical studies of rhythm and rhythmic organisation have tacitly disregarded this issue by choosing common values for parameters as tempo, note duration, and phrase length. Rhythm is certainly complex enough a phenomenon within the bounds of common music practice. However, the many times contradictory results might indicate that different mechanisms are in operation. It is therefore necessary that we take basic temporal properties of the human into account when we study rhythm. Optimal parameter values, as those common in music, are likely to obscure that, for instance, a similar experience or behaviour can be the product of different processes. Especially if the objective is to gain insights about human time in general, sub optimal values must also be applied.

The frame of thought for the following discussion is a tentative distinction between holistic and analytical processing (cf. Garner & Royer, 1968). Holistic processing is fast, passive, effortless, and phenomenally compelling, that is, yielding structure closely linked to the physical stimulus. Analytical processing can take some more time, is active, subjective to training and individual strategies, and yields structure in a more abstract sense.

Decelerating and accelerating patterns

The solid line in Figure 1 depicts a stimulus sequence with several repetitions of a decelerating pattern with continually increasing inter onset intervals (IOI). The broken line shows a less successful, but yet typical reproduction of the stimulus. The mean IOIs were 300, 450, 600, and 750 ms, and there were 2 - 11 events (period) in the pattern, which produces pattern durations (PD) from 600 ms to 8.25 s. Two groups participated, those who currently played a music instrument, and those who did not. They were in turn assigned to hearing the pattern repeated four or eight times before starting to respond. A complete report of this study is given in Madison (1997).

Results.
All groups reduced the number of intervals, such that the shortest IOI was reproduced as about half of its shortest stimulus duration, while the other intervals were lumped together. A comparison of stimulus and response was therefore not feasible, and
the periodicity was determined by applying the auto correlation function to the response sequences. The proportion response sequences with correct periodicity decreased with longer period, which is trivial because period not only conveys more levels of IOIs but also increases pattern duration.

Figure 1. A stimulus sequence with a periodicity of 10 events, and a response sequence (reproduction) with incorrect periodicity. Forty-five out of 80 responses are displayed.

The great difference between the groups in response to PD is however remarkable. There are performers who reproduce 8.25 s patterns without error, while some non-performers may consistently fail at 4 s. The difference between four and eight repetitions also interacts with music performance in an interesting way. While the non-performers benefit from the extra repetitions, the performers are hampered by it, which is interpreted to reflect the utilisation of different processes. The extra time given at eight repetitions entices the participants to analytic processing, characterised by forming and testing hypotheses, and thus forcing organisation onto the pattern. Possibly, the risk of applying an erroneous hypothesis is equal for the two groups, which leads to their similar performance at eight repetitions. With this reasoning, we would expect that holistic processing is most likely to occur for the non-performers at four repetitions. For this group, only sequences shorter than 3.25 s were correctly performed, and these included periodicities up to 8, which proves that period was not the limiting variable.

Because PD and period are correlated, and period exerted a major main effect, it was not possible to determine the relative effects of both. Another experiment systematically varied PD in six levels from 1.5 s to 8 s, which together with periods 2-11 produced mean IOIs over the pattern ranging from 136 ms to 4.033 s. Musicians and non-musicians react to short intervals - and thus short PDs - in a similar way, whereas the groups display interesting differences from 3 s PD and up. The musicians maintain ~.5 proportion correct from 3 to 8 s PD, while the non-musicians' performance declines with increasing PD. If we assume that musicians are more keen to employ strategies, that is, some kind of analytical processing, the result can be interpreted such that this can preserve structure for difficult patterns with durations as long as 8 s. Confined to holistic
processing, the non-musicians can only deal with PDs up to 3 s, which is the limit also found in the former experiment.

**Binary patterns with 2:1 ratio**

These patterns were based on possible perturbations of binary events, with one event representing a short interval, and the other an interval with twice the duration of the short. Twenty patterns were selected that formed a matrix of combinations of 8-12 events and 3-6 groups, defined as the number of long intervals in the pattern (Table 1).

Table 1  
**The patterns used, notated as long (L) and short (S) intervals.**

<table>
<thead>
<tr>
<th>Events</th>
<th>Groups (number of long intervals)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>SLSLSSLL 1 SLSLSSLL 2 SLSSLLLL 3 SLLLLLL</td>
</tr>
<tr>
<td>9</td>
<td>SSSSLLLL 5 SLSLSSLL 6 SSSSLLLL 7 SLLLLLL</td>
</tr>
<tr>
<td>10</td>
<td>SSSSLLLL 9 SLSLSSSLL 10 SLSLSSSLL 11 SLLLSLLLL</td>
</tr>
<tr>
<td>11</td>
<td>SLSLSSSLLL 13 SLLSSSSSLLL 14 SLLSSSSSLLL 15 SLLLLSSSLLL</td>
</tr>
<tr>
<td>12</td>
<td>SLLSSSSSSL 17 SLLLLLLLSSSSSSL 18 SLLLLSSSSSSL</td>
</tr>
</tbody>
</table>

* Pattern index for references in Figures 2 and 3

In each trial one of the 20 patterns was played in a repeated fashion, like a continuous loop. The participants tasks were to a) listen until the rhythmic structure was perceived, then b) tap a drum pad with a stick every time the pattern was felt to restart (detection), and, on a signal, c) tap along with every sound (tracking). Three men and ten women without formal music training performed all patterns at 3, 4, 6 and 8 s pattern duration (PD).

**Results.**

A response was considered correct if a) it occurred less than 150 ms before and less than 100 ms after the stimulus, and b) it succeeded the previous tap with the number of events in the pattern during detection, or with one event during tracking. It is thus assumed to measure the participants' ability to identify and follow the structure, more or less independent of motor performance. In order to compare with other dependent measures, in which lower values reflect higher performance, the proportion incorrect responses (PIR) was chosen. Another measure of interest is the number of repetitions before responding, which is thought to reflect the subjectively perceived rather than the factual difficulty of the stimulus sequence.

The detection task required only the anticipation of one arbitrary event in the pattern. In the extreme case, the correct representation of a small group of salient events would suffice, where the first events forebode the one with which to synchronise. Successful tracking requires on the other hand a precise account of every interval in the pattern. It is therefore conceivable that the optimal representation might differ between the two tasks, at least for some patterns. If this were the case, forming a new representation after the detection phase would probably require an extra delay. It was found that pattern duration, not pattern, had the greatest effect on this delay. These results are summarised in Table 2. Our previous findings leads us to expect one type of processing to occur within a span of 3 s. The low variability in delay at 8 s (1.15) indicates that detection and tracking both utilise the same type of processing, for all patterns and participants.
Table 2
Repetitions before responding between end of detection and start of tracking as a function of pattern duration.

<table>
<thead>
<tr>
<th>Pattern duration (s)</th>
<th>IOI range (ms)</th>
<th>Delay between detection and tracking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$M^*$</td>
</tr>
<tr>
<td>3</td>
<td>166 - 250</td>
<td>1.46</td>
</tr>
<tr>
<td>4</td>
<td>222 - 332</td>
<td>2.32</td>
</tr>
<tr>
<td>6</td>
<td>332 - 500</td>
<td>0.90</td>
</tr>
<tr>
<td>8</td>
<td>444 - 666</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note. All contrasts between $M$s and $SD$s were significant except for 3$\leftrightarrow$6 s (Scheffe test, $p < .05$).

A plausible interpretation of these data is therefore that analytical processing is employed exclusively at 8 s, while holistic processing prevails at 3 s. A change in processing is most frequently required at 4 s, probably as an interaction effect with pattern.

![Distance Weighted Least Squares](image)

**Figure 2.** Proportion incorrect responses (PIR) during detection, depicted as a function of interpolations of the four PDs and the 20 patterns. Brighter color represents better performance. Higher pattern indexes represent patterns with more events (See Table 1).

A similar result emerges for proportion incorrect responses (PIR) in Figure 2, in that detection performance is worst for intermediate values of PD, and interacts with pattern ($F(57, 627) = 1.46, p = .017$). Although patterns with higher indexes comprise more events (see Table 1), performance is not worst for these. Instead, the relation is non-linear, with patterns 6-9 and 19-20 yielding the best performance, whereas patterns 1-2 and 12-16 were more difficult. This result can not be explained with the optimal values for IOIs around 600-800 ms, (Fraisse, 1982), since they occur only at 8 s (Table 2). Since higher pattern indexes are associated with greater numbers of events, the IOIs to the right in Figure 2 are shorter, for the same PD. Tracking performance, seen in Figure 3, is distinctively different in that a) pattern has a substantial non-linear effect ($F(19, 209) = 5.16, p < .001$), b) short IOIs seem to be detrimental, and c) 8 s PD yields the lowest PIR. There is no consistent effect of the number of events in the pattern (cf.
Table 1), which is somewhat surprising considering its impact on the reproduction of the decelerating patterns. Since the number of repetitions were approximately equal (4 compared with 2.5-4.5 in this experiment), the difference must be attributed to the type of pattern.

**Distance Weighted Least Squares**

![Proportion incorrect responses (PIR)](image)

<table>
<thead>
<tr>
<th>Proportion incorrect responses (PIR):</th>
</tr>
</thead>
<tbody>
<tr>
<td>above</td>
</tr>
<tr>
<td>0.302</td>
</tr>
<tr>
<td>0.280</td>
</tr>
<tr>
<td>0.259</td>
</tr>
<tr>
<td>0.238</td>
</tr>
<tr>
<td>0.216</td>
</tr>
<tr>
<td>0.195</td>
</tr>
<tr>
<td>0.174</td>
</tr>
<tr>
<td>0.152</td>
</tr>
<tr>
<td>0.131</td>
</tr>
<tr>
<td>0.110</td>
</tr>
</tbody>
</table>

**Figure 3.** Proportion incorrect responses (PIR) during tracking, depicted as a function of interpolations of the 4 PDs and the 20 patterns. Brighter colour represents better performance. Because of the different shape of the area, both category axes are reversed as compared with Figure 2. Higher pattern indexes represent patterns with more events (See Table 1).

**Binary patterns with 1.5:1 ratio**

A subset of 10 patterns from Table 1 were presented with the "musical" ratio 2:1 or the notoriously difficult ratio 1.5:1 (Essens & Povel, 1985). Although the emergent temporal patterns may be impossible to reproduce acceptably without extensive learning, PIR estimates the accuracy of the structure and is forgiving as to minute timing.

**Figure 4.** Proportion incorrect responses (PIR) for detection and tracking, as a function of pattern duration and interval ratio. Brackets indicate .95 confidence intervals.
It was expected that performance in response to the temporally more complex, and therefore more difficult patterns would be more correctly represented at 3 s PD than at 6 s PD. For both detection and tracking, 1.5:1 was less correctly represented than 2:1, but only at 6 s PD - see Figure 4. Response precedence and SD show similar patterns, which rejects the idea that better performance at 3 s would be an artefact produced by the combination of short IOIs and PIR.

In conclusion, the bimodal distributions and interactions between PD and pattern, and other factors of complexity, indicate that at least two different processes operate within a range of PDs (and, inseparably, IOIs) entertained in music. The participants in these studies were predominantly non-musicians, and they were more correct in identifying a starting point in a pattern at short PDs, a task that resembles active music listening. The 8 s PD was on the other hand superior if the task was to track every event, similar to playing an instrument. The musically experienced participants in the first two experiments were above 3 s PD much more successful than the non-experienced in reproducing patterns with little musical relevance. The general results support that we maintain holistic and analytical processing as working definitions, pending further research. Although we know little about the properties of these processes, and less still about their foundations (Pöppel, 1996), the argument made here is that differences exist, that can be highly relevant in the analysis of time and rhythm research data.

Acknowledgement

This research was in part supported by The Bank of Sweden Tercentenary Foundation through grant to Alf Gabrielsson.

References


Madison, G. (1997) Impact of rate and pattern duration on the reproduction of accelerating and decelerating sound patterns. Submitted for publication


Cognitive Knowledge of the Basic Elements of Music

Olin G. Parker
School of Music
The University of Georgia
Athens, Georgia, USA

Music is a territory paradoxically strange and familiar. It is an integral part of everyday life, accompanying the work, the travel, and the leisure activities of humans. Music is heard daily, yet when it is listened to, what is really in the music is often bewildering because of its vastness and complexity. Music, one of the most influential of the arts, can/does satisfy humankind's need of aesthetic experiences by ministering to the human's emotional, spiritual, and intellectual requirements. It is the latter of these to which this study was directed because the profundity of musical experiences is dependent upon past pedagogically directed musical experiences. The human has an insatiable desire for learning; consequently, the cognitive aspect of musical learning must be fashioned toward helping the learning in the formation of his/her perceptual abilities and skills to most enhance music's contribution to the emotional and spiritual requisites as well. In essence a person appreciates the musical stimulus as a gestalt when he/she understands how to integrate the elements of music.

The Study

The purpose of this study was twofold: first, to bring about a thorough understanding of five elements of music to a class of musically naive adults by using descriptive words in a lecture process. This helped expand the music vocabulary of the class with respect to describing the elements of music. Second, by bringing about an understanding of the use of these elements in relation to music it greatly enhanced their ability to discriminate while listening to music. This also enabled the class to develop a more aesthetic approach toward music.

Definition of Terms

The terms used in the study were those that applied to the elements of music: pitch, rhythm, melody, harmony, and tone color (timbre). They were those given in the Kamien (1980) textbook used by the class and verified by the Harvard Dictionary of Music. (See Appendix A) Since aesthetic experience was of such importance in the study the following definitions are included:

An aesthetic experience requires psychological involvement with the aesthetic stimulus, perception of interacting events within the artwork, cognition of the interplay among the events within the aesthetic stimulus and feelingful reaction thereto. (Radocy and Boyle, p. 200);

...a person's aesthetic sensitivity will be measured by how well he can integrate the elements of music into a gestalt ... that attribute possessed by an individual which, when developed according to cultural mores, customs and dictates, enables him to make quality judgments: that is, to determine that something is or is not pleasant and the degree to which it is so. (Parker, 1973)
Another term that is relevant to this study is music appreciation. Of it, Crickmore (1968) has this description: "Music appreciation is defined as the ability to distinguish between what is generally accepted by experts as good or bad music; the enjoyment of music and knowledge are distinct though usually complementary."

Thus, the case is made for continuing to seek the most effective/productive methods of educating in all genres, in this case, the perception of the elements (and their essentials) of music.

Related Literature

In a survey of literature relating to discrimination and listening skills, measurement of affective behaviors, and the educational ideas that are involved with each, the following studies contained valuable information.

Ware (1968) studied the relative effectiveness of three organization plans of music instruction in developing appreciation for music in the elementary grades. He found that I.Q., school music environment, and music aptitude were better determinants of the ability to learn to appreciate music than were any of the instructional methods. Flohr's study (1981) however, showed that short term instruction influenced children's developmental musical aptitude. Parker (1978) said that "aesthetic sensitivity (aesthetic discrimination) shares multiple causality." It is . . . most dependent on ability and I.Q. and environment to a lesser degree." Schwadron (1966) stated that "there is considerable substantiation that the aesthetic experience is not founded on universal responses in tonal material, but acquired through education."

Colwell (1970) said that "cognition is a legitimate part of the aesthetic music behavior," Radocy and Boyle (1988) said that cognition was part of and necessary to an aesthetic experience," and Reimer (1970) stated that an aesthetic experience must include an understanding of "melody, harmony, rhythm, tone color, texture, and form." Reimer further suggests that teaching people to listen more actively by developing their perceptual abilities can increase their aesthetic responses. While performing music may allow a more personal involvement, listening is the primary way that most people in our culture interact with music. In 1980, Heddon studied the development of music listening skills. In his inquiry he examined more than 25 related studies and found that in almost all there was a significant correlation between knowledge and listening skill scores.

Hypotheses

H01 -- There will be no difference between the pre-test scores and the post-test scores on the Descriptive Word Test.

H02 -- There will be no difference between the pre-test scores and the post-test on the Indiana-Oregon Discrimination Test.

H03 -- There will be no difference between the pre-test scores and the post-test on the elements of music section of the Indiana-Oregon Discrimination Test.

Procedure

The study was conducted with the existing class of 96 adult students ages 18-35 in a music appreciation class at a southeastern university in the United States. The study covered only the elements of music and their application.
while studying music of the classical and romantic periods. Composers studied in the course were Haydn, Mozart, Beethoven, Schubert, Schumann, and Chopin. The study measured the cognitive knowledge of the elements and the subjects' abilities to use those elements in discriminating listening skills.

Method

Materials used included the answer form for the Descriptive Words Tests, the answer forms for the Indiana-Oregon Discrimination Test, the cassette tape and manual of the standardized test, and the textbook used in the class, the recordings that accompany the text (Music, An Appreciation by Kamien), and lesson plans. The Indiana-Oregon Discrimination Test (Long, 1968) is a standardized test in which the subject is asked to select the "better" of two musical renditions and indicate whether it was better in terms of melody, harmony, or rhythm. The tests were all conducted during regular class sessions.

At the first class session the pre-test on descriptive words for the elements of music was given at the beginning of the class. The class then listed as many descriptive words for pitch, rhythm, melody, tone color, and harmony as they could. The tests were collected and a lecture using audio and visual examples was given. The lecture last circa an hour. The same test was then given again. Theses scores were tabulated and recorded.

At the beginning of the next class session the Indiana-Indiana Test was given. The standardized test played tow examples of a music selection; one was correct and one was wrong, having been altered in the melody, rhythm or hamon, as previously explained. Subsequent lectures (over a period of ten weeks) covered history and/or the music of the classical and romantic periods. Listening experiences, with accompanying guides to follow, were a major part of the regular classes. Use of the five basic elements was correlated with each listening assignment.

Statistical Treatment

The pre-test and post-test scores for the Descriptive Word Test were compared using the Chi-Square test for dependent samples. The pre-test and post-test scores from the Indiana-Oregon Discrimination test were compared in two parts. The scores from the total test were compared by means of the Chi-Square test for dependent samples. The pre- and post-test scores from the second part of the Indiana-Oregon Discrimination Test were compared also; this part of the test asked which element of music (melody, harmony or rhythm) had been changed. These scores were also compared by means of the Chi-Square test for dependent samples.

Results

The results revealed that in this group of subjects the teaching procedures exceeded expectations in how much the scores between the pretests and post-tests were improved. The results from studies of this type can be very beneficial to educators.

Hypotheses H₀₁, H₀₂, and H₀₃ were all rejected. Specific data are shown in Table 1:
Table 1 | Results
---|---
1. Pre- and Post-test scores from the Descriptive Word Tests |  
2. Pre- and Post-test scores from the Indiana-Oregon Discrimination Test (IODT) |  
3. Pre- and Post-test scores from the IODT music elements |  

<table>
<thead>
<tr>
<th>Correct Scores</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. KDWT</td>
<td>61</td>
<td>131</td>
<td>$\chi^2_{25.52} &lt; P .01$</td>
</tr>
<tr>
<td>2. IODT</td>
<td>341</td>
<td>441</td>
<td>$\chi^2_{12.78} &lt; P .001$</td>
</tr>
<tr>
<td>3. IODT (elements)</td>
<td>159</td>
<td>207</td>
<td>$\chi^2_{6.29} &lt; P .01$</td>
</tr>
</tbody>
</table>

Discussion

Again, the results from studies of this type can be very beneficial to educators. When music teachers know what produces better results in listening and discriminating skills, they can plan better plan what to place into the curriculum. Although this study is too small to induce to the general population, the results show that further research would be advantageous.

The significant differences shown in the scores indicate that a planned, well-constructed listening program which emphasizes the basic elements of music will result in improvement of listening skills. The important factor for the music educators is that a structured listening program in productive. This can be done at all levels of student participation; it has significance in planning for early childhood listening experiences; it can revitalize the middle school general music program; it can add a new dimension to high school and college/university required "music appreciation" classes; and, it can be a real community service in the field of continuing education.

Recommendations for future study of the perspective aspects of learning music should include, as Sloboda (1992) suggest, "open-ended empirical investigations" of the emotions and how they interact with the cognitive knowledge (levels) of the musical elements.
References


Music Performance: Psychological and Social Aspects
Introduction

This study explored specific aspects of verbal and non-verbal auditory life among musicians and non-musicians. It is proposed that this variable differentiates between people on the grounds that adherence to an auditory style may have brought about different psychological orientations and developmental processes. The concept of differences among individuals regarding aspects of their internal auditory life was first proposed by psychoanalysts such as Nass (1971) and Noy (1968) who outlined a cognitive style based on the auditory channel originating from the preverbal infantile stage, whereby the auditory apparatus becomes a primary sensory mode, and leads to an overall auditory orientation. These theories were based, at least in part, on models of the neurophysiologist Charcot who believed that individuals could be differentiated by their sensory preferences, labelling persons as "visuels," "moteurs," and "auditifs" (Freud, 1901). However, while psychoanalytic literature may rest on intuitive and interpretative conceptualisations, it is acknowledged that these theories were not developed through a tradition of empirical rigour. On the other hand, the current research premise outlined above has re-surfaced from a different area which touches on internal auditory life - empirical inquiries exploring human foetal existence.

Recent intra-utero research methods (Hepper, 1991; Hepper & Shahidullah, 1992; Lecanuet, 1993; 1996), and infant perception-cognition research techniques (Fassbender, 1993b; 1996; Melen, 1994), offer evidence that the human neonate is predominantly an aural animal. Humans enter the world with specific perceptual acoustic experience from previous intrauterine life, and without a doubt, this prenatal climate rich in auditory experiences influences postnatal human development. For example, it has long since been observed that new-borns exhibit an auditory preference for the human voice, which suggests that prenatal auditory experiences (which contribute to such preferences) may be an unparalleled contribution to postnatal speech perception and development.

Further, many studies (Fassbender, 1993a; 1996; Gellrich, 1993; H. Papousek, 1996; Wilkin, 1993) point to prenatal experiences as the origin of human musicality and ability. Studies continue to demonstrate the fact that new-borns are already sensitive to volume, pitch, harmonic spectrum, and the duration of sound, while infants are capable of processing time-related information and differentiate sounds in terms of duration, pause length, tempo, and relative timing of rhythmic sequences (Pouthas, 1996). In an extensive review of the main research-driven theories on prenatal auditory development, Parnicutt (1987; 1993) concluded that the origins of rhythm, melody, harmony, and emotional musical meaning are associated to prenatal correlates, and underlined two common notions: (1) as a result of the wealth of acoustic learning that has taken place while in utero, humans command an auditory system that is greatly advanced already at birth; and (2) that sounds heard before birth clearly influence specific postnatal human behaviours.

But, while the concept referred to as "biological predisposition for musicality" may infer something about evolutionary requirements of the species and the adaptive significance of
human musicality (H. Papousek, 1996; M. Papousek, 1993; 1996), if children are not sufficiently engaged in meaningful auditory and musical interchanges they may undergo a process of 're-wiring' (H. Papousek, 1993). During the normal course of formal cognitive development, as children learn to pay increasing (and perhaps at times exclusive) attention to the space they see (the landscape), the space they hear (the soundscape) becomes muted through impoverished sensitivities. As a result, innate biological predispositions for audition may follow a gradually diminishing course in a visually dominant society. While the majority of children tend not to rely on the auditory mode but rather their visual senses for input and orientation, others might rely on the auditory mode to a greater extent, or perhaps find that they have cultivated a cognitive style that is of equal balance. The aim of the current study was to format these developmental speculations about neonatal predisposition which hint at the conceptualisation of Auditivity, into specific behavioural and attitudinal outcomes in order to assess the existence of a cognitive style and psychological orientation based on sensory preferences for audition.

**Method**

Two hundred fifty-four subjects (N=254) participated in the study; professional orchestra musicians (n=156), and fully employed non-musician controls (n=98). The average age of the subjects was thirty-seven years old (range=18-71), with an almost equal number of males and females, whereby the majority (66%) matriculated at an undergraduate level. Most of the musicians (83%) were players from six contract orchestras in Northwest England within string (63%) and woodwind-brass (28%) sections. The control subjects, also residents of Northwest England, represented the major sectors and vocations of the workforce, including: business-industry (41%), administration-education (17%), and legal-medical professions (12%).

Musicians were contacted through initial letters sent to orchestra halls, followed by questionnaires sent in-bulk three months later. The non-musician controls were recruited through initial individual contacts with the researcher, whereby each subject was enlisted to recruit an additional three or four controls - known as 'chain letter recruitment' or snowball sampling. Questionnaires were labelled with PIN numbers and coded (1=musicians, 2=controls) in advance; questionnaires were packaged at random and thus orchestras were not identifiable through PIN series allocation. It is difficult to assess an exact response rate because several orchestra managements reported misplaced, stolen, lost, or accidentally discarded questionnaires. In addition, there was no way to monitor exactly how many questionnaires were actually handed out by the controls. Nevertheless, if responses of musicians (156/500) and non-musicians (98/300) were estimated, these would represent a 32% response rate as the most pessimistic estimate. All questionnaire were returned by pre-addressed FREEPOST envelopes.

**Measures**

The investigation involved the development of a psychometric survey instrument by Brodsky and Sloboda in 1993 known as the Keele Assessment of Auditory Style (KAAS). KAAS is self-administered in twenty-five minutes, and is designed to elicit self-report information about developmental auditory life as a psychological orientation and cognitive style. The questionnaire survey consists of 78-items and uses a 5-point Likert Scale rating responses from never to always; an additional option is available to specify "0" for items that are not relevant, or those the respondent can't remember, or doesn't know. Sound-related items of a
general nature (items 1-55) are presented according to four developmental ages: infants (ages <5), youngsters (ages 5-12), adolescents (ages 13-18) and adults (ages >18). In addition, items about specific prior experiences involving music activity (items 56-78) are presented according to three critical periods of music involvement: childhood (ages 4-12), adolescence (ages 13-18); and adulthood (ages >18). The items link audition to one or more dimensions of human development, including: learning, memory, motivation, communication, language acquisition, imagination, inner-fantasy, intrapsychic sensitivity to sound, object relations, intimacy, experiencing and expressing affect, perception of self, interactions with others or the external environment, and self-esteem. A few examples of the more general items are...

As a Youngster...
- I felt as if I could experience my emotions far better through music than through speech.
- I preferred lullabies and songs at night time above stories and book reading.
- I was aware that I felt and heard things in music that I could not articulate verbally.
- I was comforted by a familiar tune or melody when sad.

As an Adolescent...
- I felt more comfortable within musical social settings such as ensemble and choir than other social groups.
- I identified myself as having a special gift for music.
- I imagined music in my mind when bored.
- I was aware of the similarities and/or differences in intonation (tone qualities) between my voice and my parents' voices.

As an Adult...
- During conversation I find myself listening or attending to the other person's speech patterns and vocal inflections more than the actual content itself.
- The external sound environment (soundscape) is a major factor for me when choosing neighbourhoods where to live.
- Voice tone is a major influence on whether I am attracted to someone.

Pilot-tests (Brodsky, 1995; Brodsky, Sloboda & Waterman, 1994) demonstrated that a subset of the fifty-five general items produced a 38-item Auditivity (sub)Scale which was stable and reliable (Cronbach's Alpha=0.8666; Standardised Item Alpha=0.8718), and provides a scale total score. Further, and although not discussed in this paper, post-hoc Principal Components Analysis (utilising a Varimax Rotation, Eigen Values, and Scree Plots) indicated a four-factor design consisting of thirty-three out of the original thirty-eight items (which demonstrated loadings >0.4), to be the most suitable model to accurately describe the data set; whereby the first factor accounted for 20.4% of the total variance.

Results
The study found that general orientations based on sensory preferences for audition, referred to as Auditivity, among a large heterogeneous population comprised of musicians and non-musicians was widely distributed. The item mean score of the Auditivity Scale was spread between 1.66-4.42, whereby the average item mean score was \( \bar{x}=3.08 \) (sd=0.471). Further, the distribution of Auditivity Scale Total Scores (which are summative of all thirty-eight items) ranged between 63-168, whereby the mean total score was \( \bar{x}=117 \) (sd=17.89). Between-groups analyses revealed large overlapping areas common to both groups indicating that Auditivity is not necessarily synonymous to or a precursor of musicality. That is, not all of those demonstrating high auditivity scores were musicians, nor did all musicians demonstrate high scores. Nevertheless, statistically significant differences were found between the subgroups' mean total scale scores (musicians=122, sd=17.75; non-musicians=110, sd=15.90; t=5.14, df=252, p<.001).
Further, to explore individual differences among musicians concerning psychological/musical development and orientation as based on levels of Auditivity, a median split was applied. Strictly for comparative reasons, and not assuming diagnostic criteria, musicians were classed as belonging to either the lower half (<50%) or higher half (>50%) of the sample in relation to the distribution of Auditivity Scale Total Scores (range=66-168; median=117). It is interesting to note that in spite of the fact that no significant differences were found between <50% musicians and >50% musicians regarding their general descriptive nature including age, gender, highest level of attained education, orchestra membership, orchestra section assignment, number of students per week, or weekly schedules involving amounts of practice, rehearsal and performance-related engagements, a within-group analyses found that the >50% musicians reported a variety of behaviours from childhood throughout adulthood which were significantly different than those reported by the <50% musicians. This finding compliments an earlier study (Brodsky, Sloboda & Waterman, 1994) which initially demonstrated the existence of significant differences between musicians concerning levels of anxiety as related to Auditivity. Accordingly, musicians scoring higher Auditivity Scale Total Scores were not only more "at-risk," but in fact suffered from Career Stress, Music Performance Anxiety, and Stage Fright significantly more (and to a greater intensity) than did those musicians scoring lower Auditivity Scale Total Scores. Based on the responses of KAAS items about specific prior experiences involving musical activity (items 56-78 which were not part of the general Auditivity Scale), the >50% musicians scored higher than the <50% musicians on twenty-two out of twenty-three items, and these differences were statistically significant on twelve items:

- As children involved in musical activity, the >50% musicians reported that they had more often spent time producing expressive sounds on their instruments (t=-2.50, df=154, p=.014), but only rarely attempted to imitate the sounds of nature on their instrument (t=-2.35, df=154, p=.020). Further they more often had a certain "feeling of power" when producing some types of sounds (t=2.60, df=154, p=.010), as well as a "feeling of comfort" while producing other sounds (t=-3.64, df=154, p<.001). In addition, as children these musicians attempted to imitate other performers and liked to show off with their instrument more. Finally, they more often had feelings of being different from other children who were not learning an instrument.

- As adolescents involved in musical activity, the >50% musicians sometimes felt that their parents' approval was dependent on their success as a musician (t=2.27, df=154, p=.024), and that the degree to which they mastered a piece of music influenced the degree to which they felt some control over their own lives (t=-2.87, df=154, p=.005). As teens, when faced with emotional states involving sadness, loneliness, or apprehension, these musicians more often played their instrument (t=-3.87, df=154, p<.001) as a means of coping. They more often felt not inclined to become involved with other pursuits because of their intense interest in music (t=-3.12, df=154, p=.002), and that the reason they were never really motivated to overcome the frustrations of peer interactions or academic challenges was because of their ability to escape into their own private world of music and sound making (t=-4.16, df=154, p<.001). As adolescents, the >50% musicians identified more with ensemble players than solo performers.
• As adults involved in musical activity, the >50% musicians *sometimes* felt that by playing their instrument they reconnected with the "child" in themselves \((t=-2.35, \text{df}=154, p=.020)\), and that their understanding of another person is *always* heightened if they perform music together \((t=-4.50, \text{df}=154, p<.001)\). These musicians more *often* heard sounds in their "mind's ear" when they imagined notation, and felt more psychologically lost when their instrument was masked during full orchestra *tutti* sections.

**Discussion**

The aim of the study was to explore intuitive theoretic speculations about auditory cognitive orientations and styles, which in light of more recent empirical research-based findings that support evidence of innate human predisposition for sound already at birth, take on a fresh meaning concerning individual differences based on sensory preference for audition. *A priori* acceptance of assumptions about biological predisposition for musicality, predetermined that the study measure levels of *Auditivity along a continuum* (as opposed to *auditive versus non-auditive*). The major finding of the study then, is that Auditivity - a proposed cognitive style and psychological orientation based on sensory preferences for audition - was normally distributed among the general population. This is more than ample evidence that KAAS is not measuring musical ability; some non-musicians of the general public demonstrated the same scores as did professional musicians. Perhaps this is further evidence of human predispositions for sound; while some who retain this predisposition go on to become musicians and even world class performers, others who retain this predisposition go on to become physicians and academics. Nevertheless, and although there is a great area of overlap, as a group musicians clearly demonstrated significantly higher levels of Auditivity than non-musicians.

As the operational definition of Auditivity was based on a stable and reliable 38-item KAAS subscale (derived from fifty-five KAAS items of a general nature), the study took upon itself an exploratory comparison between musicians regarding the remaining twenty-three KAAS items (whose content was specifically oriented towards musical involvement). Comparing between low(er) versus high(er) Auditivity musicians by way of a split-halves analysis, the study found no significant differences between musicians regarding descriptive information, or occupational practices. However, differences between the musician groups were seen on almost all \((22:23)\) of the music-specific KAAS items, whereby on more than half \((12:23)\) of the items these differences were statistically significant.

The weakness of the study rests on developing KAAS as questionnaire survey constructed to explore and assess Auditivity by way of gathering retrospective self-report data; the reliability of information based on biographical memory must be questioned. Further, there is every possibility musicians answered items according to certain self-beliefs and biases whereby they link sensory performance for audition to higher levels of musical ability. Finally, as PCA only indicated a small percentage of the variance accountable to KAAS factors, issues about psychometric properties and construct validity must be raised. Nevertheless, the study demonstrates that sensory preference for audition might be an important yet often overlooked factor of human development, especially where cognitive styles are concerned. Most specifically, where this dimension might be a component of musicianship and motivator towards a music career, the implications of the study are most significant to parents of children, music psychologists, music educators, music conservatories, and least not to musicians themselves.
References


Acknowledgement

The author wishes to thank ESCOM President Professor John A. Sloboda, Director of the Unit for the Study of Musical Skill and Development in the Psychology Department at Keele University, for his assistance and supervision of the current study, and for his role in developing the Keele Assessment of Auditory Style. In addition, much help was received from Dr. Mitch Waterman, of the Psychology Department at Leeds University, for his advice on matters concerning psychometrics and statistics.
The Effects of Whole-Body Acoustic Stimulation
on Subjective Relaxation, Verbalisation, and Visual Imagery
among Professional Orchestra Musicians

WARREN BRODSKY, BEN GURION UNIVERSITY OF THE NEGEV, BEER-SHEVA, ISRAEL
JOHN A. SLOBODA, KEELE UNIVERSITY, STAFFORDSHIRE, ENGLAND.

Introduction
While engaged in an intervention study evaluating the efficacy of a course on stress management for professional musicians (Brodsky, 1995; Brodsky & Sloboda, 1997), specific effects of whole-body acoustic stimulation were observed. It should be pointed out that although previous studies have attempted to document human vibrotactile sensitivities (for a review see Verrillo, 1992), as yet, there has been no systematic study of acoustic stimulation involving the whole body. Nevertheless, many researchers (for example, Standley, 1991; Madsen et al, 1991) frequently utilise a device to present auditory and vibrotactile stimulation known as the Somatron® Acoustic Massage™ System. It has been common for subjects to report this tactile-musical experience as a listening experience that is relaxing yet stimulating. Accordingly, Madsen et al conclude that "vibrotactile use in all of its various ramifications merits careful attention. Aspects relating to whole body... relaxation or stimulation caused by various frequencies without, or in combination with, various musical selections are all fertile areas for investigation" (pg. 21). Clearly, while it has been known for some time that music vibrations penetrate the body, it is essential that current studies focus on more specific effects through cautious research. During our clinical trial of a therapeutic treatment for musicians utilising the Somatron Recliner, we observed qualitative differences among responses of the sample regarding subjective relaxation, verbalisation, and visual imagery. We have every reason to believe that these differences are solely attributed to whole body acoustic stimulation, that is, the application of music generated vibration to cutaneous tissue involving large surface areas.

Method
Fifty-four (N=54) professional symphony orchestra musicians participated in the study. The majority (68.5%) were string players, with all other major orchestral instruments being represented by a minimum of three players. The average age was thirty-six (range=22-55 years), with an almost equal number of males and females. Forty-eight of the participants (89%) had completed formal music training at the undergraduate degree or diploma level.

Musician subjects were recruited via a questionnaire sent to orchestra managements. Four research sites were initiated in three cities across North West England. Musicians of each site were assigned a PIN number and listed according to descriptive criteria such as gender, age, orchestra section, principal instrument, and status (contract/freelance) - these served the random assignment to intervention conditions. The procedure was two-fold: (a) three subjects were matched according to descriptive criteria; and (b) one of the three matched subjects was randomly assigned to one of the three experimental conditions. Musician-subjects were both matched within each site, as well as between sites. The three experimental conditions were:

1. Verbal. Seated in the recliner, the subjects received individual contact consisting of verbal conversation, visual imagery tasks, and relaxation exercises. This condition was identified as the control group.
2. Music. Seated in the recliner, the subjects received individual contact consisting of verbal conversation, visual imagery tasks, and relaxation exercises supplemented with pre-recorded music from commercially available cassette tapes and/or CDs presented through headrest-mounted speakers.

3. Somatron. Seated in the recliner, the subjects received individual contact consisting of verbal conversation, visual imagery tasks, and relaxation exercises, accompanied by pre-recorded music from commercially available cassette tapes and/or CDs presented through speakers embedded in the recliner at various body locations.

Each musician received eight individual 50-minute sessions over eight weeks. Every session opened and closed by completing the Profile of Mood States (POMS) Monopolar Form (McNair, Lorr, & Droppleman, 1971). The POMS was used as a method to monitor effects of emotional stimulation within sessions, as well as across the time period of eight sessions. Self-administered in roughly five minutes, whereby respondents rate sixty-five adjectives on a 5-point scale, POMS measures six affective states and a global total score. After completion of the POMS, relaxation exercises were implemented. Then, verbal conversations focused on either cognitive appraisals or career development, or focused on theories connected to action tendencies. Visual imagery tasks were implemented as means to widen self-awareness. The content of the conversations, cognitive appraisals, and imagery tasks were simultaneously transcribed throughout each session.

Results
A repeated measures analysis of variance (ANOVA) indicated statistically significant positive general effects of time (i.e. across eight sessions) for all musicians regardless of the experimental condition on four out of six POMS mood states. These reflected a progressive reduction in tension-anxiety \( [F(1,7) = 9.56, p < .001] \), depression-dejection \( [F(1,7) = 2.56, p < .014] \), anger-hostility \( [F(1,7) = 3.14, p < .003] \), and confusion-bewilderment \( [F(1,7) = 9.43, p < .001] \). Further, a reduction of the total scale score was demonstrated \( [F(1,7) = 2.91, p < .006] \). However, no significant score differences were observed between the experimental conditions, that is, no interactions between experimental condition \( \times \) time were demonstrated.

Nevertheless, specific qualitative differences were observed between both experimental music conditions (Music and Somatron) in comparison to the no-music control (Verbal) condition, as well as between the two music conditions themselves. These between-group differences were assessed through a content analysis of session transcripts. The analysis attempted to account for the qualitative nature in which the subjects described: (1) how they felt after relaxation; (2) the intensity and depth of self-disclosure during narratives and cognitive appraisals offered verbally after relaxation; and (3) a count of occurrences of visual scenes during imagery tasks offered verbally after imagery exercises had concluded.

The transcripts presented herein represent a random sampling of the total four-hundred thirty-two hours subject contact. While space limitations allow for only a few examples to illustrate specific points, the reader can be assured that the selection process was blind to descriptive information, and do in fact reflect the nature and character of the full sample. It should be noted that each vignette is prefaced by a case number, and that for the sake of anonymity all references to principal instrument, principal instrument, secondary-instrument, as well as names of cities, concert halls, and music academies have been genericised using the labels piano, violin, cello, London, Theatre, and Music College - all of which are presented in italics. Although the sessions were not tape-recorded, every attempt was made to represent subjects' comments and narratives as original as possible including grammatical errors and use of slang. Consequently, where additional explanations were necessary, these appear between squared brackets ([ ]).
1. Subjective Relaxation. In general, the majority of musicians in the Verbal condition did not object if relaxation exercises were skipped and a session began with verbal conversation, whereas the majority of musicians in both Music and Somatron conditions disapproved if sessions did not begin with music-accompanied relaxation. It is further interesting to note that transcripts of subjects in the Verbal condition made little reference to the recliner or other aspects of the experimental environment or experience, while transcripts of Music subjects referred to several variables including the recliner, repertoire, listening behaviours, and audio quality. However, the transcripts of Somatron subjects are not only filled with statements describing their perception of the experience, or refer to the recliner, levels of relaxation, or repertoire, but highlight the affective quality of the music vibrations. It should be pointed out that verbal conversations did not take place during relaxation exercises, but following it.

The majority of the musicians in the Verbal condition referred to levels of relaxation by stating: Felt really relaxed. or Feel more relaxed now [after relaxation] than when I came in. Some commented about how their breathing patterns fluctuated throughout the exercises, while others commented how they suddenly became aware of their body. Several subjects commented about the inter-relationship between relaxation and performance.

At first I couldn't seem to relax my thigh muscles. [Makes little physical effort to tense his muscles. After the exercise, he says.] I feel more relaxed now than when I came in... this would be the best way to perform - alert but not tense!!

The majority of musicians in the Music condition referred to the audio quality or repertoire by stating: Don't like that piece. Didn't fancy that orchestra. or What a naff arrangement. While some felt that the music experience was great, others expressed that music might be detrimental to relaxation.

The majority of musicians in the Somatron condition referred to levels of relaxation, pleasure/displeasure of vibration, interference/facility of vibration on relaxation, repertoire, comparison of the audio quality to their home stereo, and comparison of this experience to those involving the stage. However, what stands out clearly is the emotional quality of the Somatron experience - sometimes catching them off-guard flooding them with unanticipated feelings, while at other times elevating them to a peak.

325
2. Cognitive Appraisal. Musicians were asked to evaluate the meaning and impact music had had in their lives prior to and during the process of becoming a musician; these were relayed in the form of an autobiographical narrative. Previously, Sloboda (1989; 1991) found that autobiographical memories of music were particularly retrievable tapping musical experiences covering a wide range. He concluded that this method was an effective means to recall significant emotional responses to music experiences. However, unlike Sloboda's previous report, our subjects' narratives reflected one of two distinctive styles which are qualitatively different: surface-narratives versus depth-narratives. While some musicians offered narratives that remained somewhat on the cognitive surface (describing their biographical history as a mere sequence of events), others went into more emotional detail under the surface (and often touched affective lines which were emotionally painful). In general, surface narratives were shorter texts reflecting on average 70 words (sd=35.07, range=37-117 words), while the depth narratives were more than twice as long reflecting on average 230 words (sd=69.80, range 261-414 words). It is interesting to note that while only 26% (or 5:19) of the subjects in the Verbal condition offered depth-narratives, 63% (or 10:16) of subjects in the Music condition and 58% (or 11:19) of subjects in the Somatron condition offered depth narratives. An example of each type follows.

375 It is an expression of self. More than expressiveness. It gave me a discipline - had to be done every day. It became
an obsession. It's a focus of life. It is almost a religion.

032 I personally feel that wrongly through my life music has been a false G-d. I have given it far too much value. I thought
it was the be all end all to everything else. It gave me a feeling, a personality, a purpose to strive for. Value and
importance.. kudos At 3-4 I began piano, my mother was my piano teacher. Both my parents are musicians and
travel as performing musicians. It was the 'kiss of death' so to speak. Mom and I didn't get on as teacher and pupil..
I was difficult to teach. At 8 I found a violinist who was wonderful in London. This teacher didn't talk down to me.
From the beginning I identified myself as a violinist. By the age of 10-11 I had a strong feeling of destiny. I began
practising from about 11 years old. The instrument gave me a great feeling. To try things out on the violin. I was
enamoured by the repertoire. The human speaking quality. You can talk with it, or communicate through it. My twin
brother was more of an extrovert than I was. Unlike me he was the life of the party. My music compensated in this
rivalry. I spent much time in Youth Orchestras. Saturday workshops from 12-16 years old. At 17 I joined the Music
College. Most of my friends are musicians (though about 5% are non-musicians). I was a late bloomer (sexually) and
all my intimate contacts really only occurred from and after college. Thus, all my intimate relationships have always
been with musicians. I feel that the reason for this is not just circumstance [being among musicians] but
understanding. [He feels a specific dynamic relationship to the instrument.] A total togetherness. Tonal
makeup, variety, big, resonant, rich. The instrument is the sound in me. I feel closer to the instrument than to
humans. The way I and the violin sound together leaves me feeling self sufficient, needing no further human contact or
relationships (only seldom do I need others).

3. Visual Imagery. Imagery tasks were implemented as a means to widen self-awareness. Conversations did not take place during imagery tasks, but following them. A content analysis of reported imagery demonstrated many commonalities regardless of condition. For example, in one task involving an imagined stage-performance, the majority of musicians were able to visualise a familiar venue complete with audience. Further, most subjects who visualised themselves performing on stage, watched themselves perform from either a location in the wings (i.e., off-stage to the right or left), or from behind the orchestra. However, and in addition to these, only musicians from the Music and Somatron conditions reported feeling sensations as if they were actually performing on stage.

294 The venue was The Theatre. I could see the faces of the audience, but they were still a general mass of people.
There were identifiable members of the audience that I personalised such as my parents, daughter, neighbours, some
colleagues, and a few players. They seemed to be excited. I couldn't see myself as a performer play... but I could feel
as if I was playing (felt realistic). I could feel the technical aspects of the performance. I felt as if I was playing well, not nervous, however slightly more so when I thought about my parents being in the audience. In general it was a pleasurable experience.

While visual imagery tasks involve projection and introspection, supplementing these with music and/or music-generated vibration seems to both heighten the experience, as well as evoke many more memories and associations. For example, in one task where subjects from both experimental conditions were asked to perceive the heard music as a television movie soundtrack, and visualise the teleplay as vividly as possible, the majority of musicians reported: (a) images of nature and the great outdoors including scenes on a beach, in the hills, or near a forest; (b) images involving motion or travel including cars, boats, or on horseback; and (c) images involving figures (some of whom were recognised as neighbours or relatives). Nevertheless, a content analysis of this reported imagery demonstrated interesting differences between subjects - albeit this time between the two experimental music conditions themselves. In general, the imagery reported by subjects of the Music condition were shorter texts reflecting on average 57 words (sd=19.20, range=36-90 words), while the imagery reported by subjects of the Somatron condition were a bit longer reflecting on average 81 words (sd=24.91, range=54-126 words). However, the main difference between these two groups surfaced by counting the occurrences of visual scenes as reported in their imagery; these reports were all offered verbally after imagery exercises had concluded. The analysis revealed that 69% (or 11:16) of the subjects in the Music condition reflected monothematic or one-scene imagery, while 79% (or 15:19) of subjects in the Somatron condition reflected multiple-theme/scene imagery. An example of each type follows.

285 Driving along the road, two people male and female, on a warm day. The car was a convertible 60's type with big chrome wings. In a national park, no trees, desert area.

387 A couple in a rooftop restaurant by the window with the city streets below. Feeling relaxed... enjoys the meal. The waiter told the man of a phone call for him, after which he came back white-faced and crying. It began to rain and pour. The window blurred. Food and drinks now looked as if it was awful. Scene of a park. Same couple but he was less defined. Girl was really happy hopping along the wall. She climbed a tree. Expressing happiness by screaming happily. Still not enough happiness and then before flying about like a bird while the man just sat in the park ignoring her. She laid down in the field and fell asleep.

**Discussion**

Although our primary research aim was not to explore the qualitative effects of whole-body acoustic stimulation, musicians' responses to this stimulation was of a very marked character. It must be pointed out that the transcription method used may be less than totally reliable particularly as it may be open to experimenter bias. A more valid procedure would most certainly have been to audio-tape or video-record the sessions that would thereafter undergo a process of transcription and assessment by individuals blind to experimental conditions. Nevertheless, since these effects were unexpected and unpredicted by the researchers, we may be somewhat confident that these effects are real, not artificial.

Other recent studies suggest that the effects of vibrotactile stimulation are subtle and indirect. For example, Walters (1996) reported using the Somatron to investigate the effects of vibrotactile stimulation on thirty-nine women awaiting scheduled gynaecological surgery. Randomly assigning subjects to one of three groups (vibrotactile, music only, and no-treatment control group), Walters found no significant differences between the groups regarding standardised post-operative assessments. However, in comparison to control
subjects, women participating in either music condition spent significantly less time in surgery and post-anaesthesia care, as well as received significantly less post-operative medication. Further, subjects receiving pre-intervention vibrotactile stimulation demonstrated the least systolic/diastolic blood pressure fluctuation throughout the surgical experience, in addition to significantly less post-operative apprehension. Walters concluded that conditions consisting of either music-only or vibrotactile stimulation seem to be equal in effect as "the two appeared not to differ significantly from each other on any measure apart from the self-report data" (pg. 261-262).

Conceivably, the effects of vibrotactile stimulation via the Somatron® Acoustic Massage™ System are not readily amenable to standardised psychometric assessments, but rather self-report measures, verbal narratives, and cognitive appraisals. Our data suggests that vibrotactile stimulation generates a strong sensory experience which leads to emotional intrapersonal processes that produce rich and deep post-stimulation verbalisation. An explanation for how this type of stimulation leads to these specific and distinctive effects is a matter for further research.

References

Acknowledgement
The authors wish to thank Mr. Byron Eakin, President, Somatron Corporation (formerly Somasonics Inc.) in Tampa Florida for providing Keele University with the Somatron® Recliner which was used in the Clinical Trial Intervention.
For all present at a live performance of a repertoire piece, the communication and reception of the musical score are a central concern, but the performance and reception of the work are dependent on a number of complex social factors. There are socio-cultural rules which constitute and give value to the composition and its performance (e.g., the use of particular scale systems, a model aesthetic, the size of an ensemble). These rules also constrain more specific inter-performer, inter-audience and performer-audience behaviours. In addition, there are moment by moment issues of co-ordination, co-perception and feedback between performers. To successfully perform a musical work requires subtle co-ordination of rhythm, intonation, dynamics, and articulation as well as more stylistic matters as phrasing and ornamentation. It involves a complex form of social knowing in which each player takes the "meaning" which is specified at any particular point in a performance as a "means" for further specifying what is afforded by both the part before him/her and the "musical product" of the joint activity of all the players. These two layers of factors are not mutually exclusive as immediate performer-performer interactions and socio-cultural factors inevitably interact, sometimes resulting in modifications to performance practices.

Studies with a focus on social issues have included work by Young and Colman (1979) and Murningham and Conlon (1991) on string quartets, and Faulkner (1973) and Atik (1994) on orchestras. In the first case, no empirical investigation was undertaken, and in all four cases the musical ensembles were studied as "organisations" with conflict, co-operation and leadership being at the centre of the research. Although these studies identify and highlight some important social issues, the musical performance as a series of communicative actions is not considered in detail.
Two recent studies which do focus on some of the details of musical communication are those of Weeks (1996 a & b). Working within an ethnomethodological tradition, Weeks has examined the achievement of musical coordination in two settings. In the first he examined how, during a concert performance of the Saint-Saëns Septet, a set of amateur musicians successfully restored synchrony after a number of timing errors occurred. In the second he focused on some of the processes whereby concert performances of orchestral music are collectively achieved in rehearsal. In this second study he focused on a youth orchestra rehearsing a section of the first movement of Beethoven's 8th symphony in which a number of problems arose in a 'cello and bass passage. In both of these studies Weeks is using the characteristic ethnomethodological strategy of attempting to gain insight into the ongoing accomplishment of music-making through instances of its disruption. Whereas in the latter case he makes sense of what is going on through the analysis of the conductor's talk, in the former case where there is an absence of talk, musicians' "insider knowledge" is required in order to recover just what the musicians are doing.

Although research in the psychology of music has virtually overlooked the fact that performers are social agents, there have been some useful developments in social and psychological theory upon which to build. In a series of four essays the social theorist Alfred Schutz reported his phenomenological investigations of musical phenomena. In these essays Schutz is not just exploring the phenomenology of musical experience but he is using music-making to clarify the rich and complex nature of human communication. One of these essays, "Making music together", explores in detail some aspects of the sharing of meaning in music-making (Schutz, 1951). For Schutz the communication of musical meaning is a distinctive example of "pre-communicative", non-conceptual social interactions. For Schutz, all communication, including musical communication is founded upon a type of social interaction which he terms the "mutual tuning-in relationship". The communication of musical meaning is accomplished against the background of a socially conditioned stock of musical experiences. In his essay Schutz explores this "mutual tuning-in relationship" with respect to various aspects of the performer/composer/audience relationship. Schutz extends his analysis to include the relationship between two or more performers of a musical work. Each performer's action is oriented not only to the composer's intended meanings, but also to the audience and any co-performers. In a chamber music setting the immediacy and visual co-presence of the other performers make it easy for each to take into account the others' activities. Each of the musical actions is available to the others and each must be able to anticipate the activities of the others.
In so doing they are living through the flux of the musical events, following both the suggestions that the composer has made in the work itself together with the suggestions of other performers.

Clark and Brennan (1991) offer a further theoretical perspective in their analysis of "grounding" in conversation, their framework can be extended to musical communication. Grounding in musical communication would involve both the coordination of content and process. The former involves assuming a vast amount of shared information or common ground - mutual knowledge, beliefs, and assumptions. The latter involves the synchronization of entrances and exits, dynamics, changes in tempo etc. Music-making is built on such common ground. Coordination of content requires the use of the stocks of musical knowledge referred to by Schutz - a familiarity with performing traditions, musical notation practices etc. To coordinate on process requires an updating of common ground moment by moment. Clark and Brennan believe that there are two main factors that shape grounding: purpose - what the people involved are trying to accomplish in their communication; and the medium of the communication - especially the dimensions of communication that are available within the medium of communication. In their analysis Clark and Brennan identify a number of dimensions of communication by means of which such coordination of process can be accomplished. These include co-presence - whether two or more people share the same physical environment; visibility - whether the people are visible to each other; audibility whether they can hear and be heard by one another; co-temporality - whether one receives at roughly the same time as the others produce; and simultaneity - whether each can send and receive at once and simultaneously (Clark & Brennan, 1991). It should be apparent that these dimensions are clearly involved in situations of small-scale music-making but that some of the dimensions are not available to performers in orchestral music.

A final example can be drawn from a rather different domain of music-making. In a comprehensive study of jazz, Berliner (1994) provides an invaluable commentary on the social processes that are involved in the "give and take" of the improvisatory exchanges that take place in jazz.

We have chosen to study the string quartet since it is one of the most enduring musical ensembles, in existence in its current form from the mid-eighteenth century. The consistency of the string timbre and the use of essentially comparable instrumental techniques has generally led composers to give each of the four instruments similar musical elements to perform. As in other types of small group,
performer individuality has continually to be balanced with group role. This seems to provide both personal freedom and security to the individual performer.

In this as in any other field of inquiry, it is important to have a rich description of the phenomena which are to be explained. As a first step towards an understanding of communication in chamber ensembles we shall present material from a case study of a string quartet. The string quartet will be analysed playing a single movement from a repertoire piece. The focus of our analysis will be the musical communication between the players. The data will allow us insight into the musical goals of the players and the means by which they try to achieve these in performance, and how well these goals are achieved and provide information about the nature of the communication between the players. The recordings will also be shown to the performers who will provide their own accounts of both their musical goals and of the nature of the communication involved.

References:


INTRODUCTION

Performance anxiety or stage fright in musicians means their state of mind and body getting very nervous before or during public performance. This includes various levels of anxiety from a transitory light symptom to a serious problem which causes the performer great distress. According to several surveys using self-reports, over 21 to 24% of respondents indicated that they experienced marked distress while performing, and 14 to 16.5% considered that it seriously impaired their performance (Fishbein & Middlestadt, 1988; Wesner, Noyes, & Davis, 1990; Kemenade et al., 1995). Stage fright was positively correlated with neuroticism and fear of social situation, and negatively associated with extraversion (Steptoe & Fidler, 1987). A number of studies found that high levels of performance anxiety have been associated with excessive task-irrelevant thoughts that impair ability to focus on the performance, and some clinical researches have reported that cognitive therapy or cognitive-behavioural therapy were viable treatment approaches for high-anxiety musicians (Kendrick et al., 1982; Clark & Agras, 1991).

On the other hand, there are researchers who urged, from the point of view of Drive Theory, either that anxiety improved the quality of performances or that subjects with high formal training performed in a superior manner in anxious situations (Hamann, 1982; Hamann & Sobaje, 1983). In these studies, however, the State-Trait Anxiety Inventory (Spielberger et al., 1970) was used as a questionnaire to determine musicians' performance anxiety. There is little evidence that we can consider these two as one and the same. ‘A prima facie case can be made for important qualitative distinctions between the demands made on performing musicians and other forms of evaluative performance. The musician must be in absolute control of motor coordination involving the finest muscle action, has to trust memory, and at the same time must feel and project the music to sophisticated audiences with authority and conviction’ (Kendrick et al., 1982, p.354). It seems to be appropriate that a questionnaire of musical performance anxiety should deal with not only cognitive but also of psychosomatic or psychomotor features which would interfere with performance. In other study, musicians were designated as “highly-anxious” if they reported at least one episode of performance-related “tremor” while performing in public (Fredrikson & Gunnarsson, 1992). At any rate, we need a more objective scale with established psychometric properties to measure this anxiety and it is necessary therefore ‘to learn more about the nature and extent of performance anxiety among present-day musicians’ (Wesner et al., 1990, p.177). In order to do this, we distributed the Performance Anxiety Questionnaire, (PAQ; Cox & Kenardy, 1993) and some other inventories among music
students of Japan.

METHOD

Subjects: Two groups of students participated in this study. Groups I and II comprised 40 undergraduate music students each, with a mean age of 20. The students of the two groups are performance majors and they have had 14 years of formal study on the average. All students are women. Questionnaires were distributed to all of them at the Faculty of Music, Ueno Gakuen University, Tokyo.

Measures and Procedure: The PAQ designed by Cox and Kenardy was used as a measure of performance anxiety. The PAQ is a measure of how frequently the subjects experience cognitions, (it contains 10 describing cognitive feelings: e.g., I worry about my performance) and somatizations (10 describing somatizations: e.g., I feel tense in my stomach) in three performance settings: practice, group public performances, and solo public performances. Subjects respond on a five-point scale to every statement for each performance setting by imaging each performance situation.

After the PAQ, students of group I also received the State-Trait Anxiety Inventory (Spielberger et al., 1970/1991), as well as a questionnaire on their motivation in musical performance with the following instruction: ‘Why do you play music on stage? What is your purpose in performing music? Please write down the reason.’

As to the students of group II, after the PAQ, they also completed the Yatabe-Guilford Personality Inventory (Y-G test). This test is a typical personality inventory in Japan, adapted from three forms of J.P. Guilford Personality Inventory by Yatabe, et al.(1951/1965) so as to suit the Japanese cultural environment. Personality traits are measured with a 120-item questionnaire in this test. These items were selected to represent 12 personality factors: D: depression, C: cyclic tendency, I: inferiority feeling, N: nervousness, O: lack of objectivity, Co: lack of cooperativeness, Ag: lack of agreeableness, G: general activity, R: rhathymia, T: thinking extraversion, A: ascendance, S: social extraversion. It is possible to judge individual personality type (A to E) by drawing a profile of each factor’s percentile.

RESULTS

Results of group I: The PAQ score was calculated by adding each rating point (1 to 5) of 20 items in three settings. First, a cut-score of 41 based on the group median on the STAI-T was used to divide the students into two groups in the same way as Cox, et al.(1993), but there was no difference between the PAQ scores of the two groups. Then, our students were divided among low trait-anxiety (33 to 44), high trait-anxiety (45 to 54) and extreme high trait-anxiety (more than 55) groups according to the appraisal standard of STAI-T. Mean scores of PAQ among three groups are shown in Figure 1. The means indicated that students with extreme high trait-anxiety experience higher performance anxiety level compared to students with high trait or low trait-anxiety only in a practice setting (t(23)=1.74, p<.05). Regarding the score of the state anxiety, students were divided into low state-anxiety (41 or less) and high state-anxiety (more
than 42) groups. Figure 2 shows that there was significant difference between the two groups in the level of performance anxiety for solo setting (t(38)=1.53, p<.1). Our figure is similar to that obtained from the study of Cox et al.(p.55) about levels of social phobia. These data suggest some relationship among performance anxiety, state anxiety and social phobia.

As to their responses to the questionnaire about motivation in musical performance, students could be divided into four groups on the basis of their motivation as follows:

(a) achievement directed group (32.5 per cent) made statements such as: ‘I think musical performance on stage is a great chance to show my ability in music to the audience’, ‘I want to be evaluated my musical achievement by audience or judges through performance’, and ‘Musical performance at an audition, for me, is an ordeal to make progress and to succeed in music’. (b) expression-communicate directed group (22.5 per cent) made statements such as: ‘Musical performance is to express myself and to create my own world with musical sound’, ‘I want to communicate the beauty of music to an audience by my performance’, and ‘Because I feel a desire to perform this wonderful music and I want please the audience with it’. (c) self-satisfaction group (17.5 per cent) made statements such as: ‘It is a pleasure to perform music before the public, and I am always satisfied with my performance’, ‘I am very glad to be applauded my performance’. (d) unaware group (17.5 per cent) gave descriptions such as: ‘I have never thought about it’, ‘I have no reason in particular to perform because musical performing is a part of everyday life’ and ‘as I have played a piano ever since I can remember, it became just like a habit for me’.

It can be seen in Figure 3 that the achievement directed subjects had a significantly greater increase in performance anxiety (M=63.7, SD=12.71) due to solo performance setting as compared with the expression-communicate directed subjects (M=53.1, SD=5.12; t(20)=2.14, p<.05). There were no differences between the score of achievement directed group and the scores of the other two groups ((c)M=62.3, SD=9.7; (d)M=64, SD=7.6) in all settings. These results indicate that musicians who are motivated by a quite strong desire of expression or communication have less performance anxiety than musicians motivated by other reasons in musical performance.
Results of group II: Table 1 summarizes the distribution of responses to the three-point global rating of performance anxiety, ranging from a mean of 41.28±3.9 in the lowest to 64.62±3.1 in the highest group. The scores on the PAQ were significantly different across all three groups (Ryan method; High PA vs. Low PA: t=15.3, p<.001; High PA vs. Mid. PA: t= 7.8, p<.001; Mid. PA vs. Low PA: t=10.2, p<.001). Table 1 also presents percentages of Personality types diagnosed by Y-G Personality Inventory which hold in each levels of performance anxiety. The PAQ scores of each personality type in three settings are shown in Table 2. Although there were no difference among them in different settings, there was a different tendency among percentages of the five personality types held in Low PA(Chi-square=8.62, df=4, p<.10, see Figure 4). This suggests that subjects of D (Director) type, whose distinctive characteristics are being active, stable and having social adaptability, tend to have less performance anxiety compared with other types’ subjects.

Correlations between performance anxiety and personality factors were calculated for each settings. Performance anxiety was positively correlated with ‘inferiority feeling’(r=.513, p<.001), ‘depression’(r=.361, p<.05) and negatively associated with ‘general activity’(r=—.312, p<.05) only in solo setting. There were no definite associations between performance anxiety and ‘nervousness’ in our music students (r=.278, p<.10) while the correlation between this anxiety and neuroticism was clearly observed among professional musicians in the study of Steptoe et al.(1987)[Note: It is ascertained by factorial analysis that Y-G test is basically similar to Eysenck’s MPI]. Our finding with respect to the high correlation between performance anxiety and inferiority feeling suggested us the relationship between performance anxiety and perform score in the faculty. The correlation was —.264 (p<.10).

Table 1. Level of performance anxiety and associated characteristics: means with standard deviations in parentheses

<table>
<thead>
<tr>
<th>Total number(%)</th>
<th>Low PA</th>
<th>Middle PA</th>
<th>High PA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance Anxiety Questionnaire</td>
<td>13(32.5)</td>
<td>14(35)</td>
<td>13(32.5)</td>
</tr>
<tr>
<td>Setting: Practice</td>
<td>41.3(3.9)</td>
<td>54.5(3.1)</td>
<td>64.8(3.1)</td>
</tr>
<tr>
<td>Group public</td>
<td>37.7(4.1)</td>
<td>42.1(5.8)</td>
<td>51.4(8.9)</td>
</tr>
<tr>
<td>Solo public</td>
<td>36.0(6.5)</td>
<td>55.4(6.2)</td>
<td>63.3(7.9)</td>
</tr>
<tr>
<td>Type of YG Personality Inventory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type A (%)</td>
<td>2(50)</td>
<td>1(25)</td>
<td>1(25)</td>
</tr>
<tr>
<td>Type B (%)</td>
<td>0(0)</td>
<td>5(56)</td>
<td>4(44)</td>
</tr>
<tr>
<td>Type C (%)</td>
<td>2(28)</td>
<td>3(43)</td>
<td>2(28)</td>
</tr>
<tr>
<td>Type D (%)</td>
<td>9(60)</td>
<td>3(20)</td>
<td>3(20)</td>
</tr>
<tr>
<td>Type E (%)</td>
<td>0(0)</td>
<td>2(40)</td>
<td>3(60)</td>
</tr>
</tbody>
</table>

Figure 4. Performance anxiety level and personality type
DISCUSSION

Our findings show that, first, the PAQ designed by Cox & Kenardy (1993) is so useful in measuring performance anxiety because it includes not only psychological but also psychosomatic items. Assessment by the PAQ may be better than assessing performance anxiety using the State scale from the STAI since the former items seem to reflect a more real state of stage fright. Moreover, the results indicate that setting is an important determinant. Since subjects assess their anxiety in three different settings separately, they can draw out their experiences from memory easily and objectively.

Second, students with high levels of state anxiety are more anxious during solo public performance than low state anxiety students. Other studies have reported similar results with relation to performance anxiety and levels of social phobia (Steptoe & Fidler, 1987; Cox & Kenardy, 1993). This suggests that performance anxiety signifies a distress compounded with specific fears of social situation, in which a solo performer is being exclusively evaluated by the audience on various sides not only about performing. Regarding to the trait anxiety, there was no difference in two public settings whereas students categorized as having extremely high trait anxiety already exhibited significant increase of performance anxiety in practice setting.

Analysis of self-statements about motivation in musical performance revealed that thoughts or motives about performance could be grouped into four. Our third finding suggests that the motivation of the performer has an effect on his or her performance anxiety. In particular, music students who strongly hope to express themselves eloquently in music, together with a positive attitude to the audience, indicated significantly low performance anxiety. This type of motivation may be useful for these students to attend to task-oriented and positive thoughts during performing and to catch a performing occasion probably not threateningly.

The positive association between ‘inferiority feeling’ and performance anxiety is of great interest, fourthly. This factor is considered as a personality factor closely connected with the emotional instability in Y-G test, as well as ‘depression’ and ‘nervousness’. In support of this, E ("eccentric") type students, whose distinctive characteristics are being introverted, passive and unstable emotionally, show the highest performance anxiety among the five personality types (see Figure 4). It can be expected that persons having a lot of ‘inferiority feeling’ and ‘depression’ tend to arrive at a negative thinking out of an adverse performing experience. According to Steptoe & Fidler (1987) and Steptoe et al. (1995), thoughts centered on a realistic appraisal of the situation or performance quality appeared to be more adaptive than worrying thoughts on making a mistake in performance. But what causes the performer to keep the composed perception during public performance? Further research on structural links between perception and various aspects, especially that of personality and motivational aspects in musical performance is still to be done.
REFERENCES


Factors Influencing Singing Development in Poor-Pitch Singers
Berit Lidman-Magnusson
Centre for Music Pedagogy, The Royal College of Music, Stockholm, Sweden

INTRODUCTION

A defective singing ability often entails a feeling of being outside a singing community. This can be experienced as a social disability. If a child, who is a poor singer, is told early by her/his family, teachers or others that s/he cannot sing, this often results in an inhibition for life (Lidman Magnusson, 1994). There are descriptions of music teachers whose negative statements about people's singing ability probably established and strengthened a dysfunction in the singing ability (Palmgren, 1885; Lind, 1897; Reberg, 1993; Rolle, 1914; Torell, 1945). Defective singing ability is described and named in different ways (Welch 1979a). The most common criteria of a defective singing ability are:

* little ability to follow the melody contour
* little ability to follow the melody rhythm
* singing at very low pitch
* the range in singing is limited to a few tones

The range of the voice also seems to be limited to just a few notes (Lidman Magnusson, 1988, 1994). Terms used are low pitch singers, inaccurate singers, poor pitch singers, weak singers, monotones and onchi singing. I have chosen to use Dysfunction in the Singing-ability-System (DSS) to describe an undeveloped singing ability. This term stands for inhibition - cancellation, slowing down and ceasing - of several of the mental processes, which are linked to the creation of singing ability. Given this definition, a DSS- diagnosis provides a hint that certain central phases in singing development perhaps have been blocked.

My own research indicates that one can suffer from inhibited singing-development without being inhibited in a more general way, that is, solely the singing-development may be blocked or damaged. Descriptions of different kinds of undeveloped singing ability are found in a number of articles and dissertations, for instance, Norioka (1992), and Hermann (1983, cited in Bruhn, 1991).

Possible hindrances to singing development.

The problem for a "poor pitch singer" is probably due to an interruption of the singing development at a lower developmental level. The "poor pitch singer" has not continued his or her singing development as most people do. This does not mean that children who have come to a standstill in their singing development are unable to sing. Rather their singing development has not been sufficiently supported by education and suitable instruction at a certain point. Singing can be regarded as an interaction between sensory, motor and cognitive processes. Most scientists who have investigated poor pitch singing have assumed that singing disturbances arise in the motor area or are due to cognitive factors. They may also depend on hearing deficits but only in a few cases. Even deficient intelligence and difficulty with schoolwork can affect singing ability, as well as social disturbances, such as frictions in the family or in the environment (Bruhn 1991). Reasons for singing problems may also stem from emotional problems in the form of inhibitions, misfortunes, aggressiveness etc. (Bruhn, 1991). Robinson (1963, cited in Welch 1979a) found that orphan children who were disturbed and unhappy were without exception classified as "poor pitch singers". Several studies clearly show the importance of an favourable home environment and of the family's positive attitudes to singing development and artistic talent in children. The singing ability of a grown-up possibly depends on an interaction between different factors among the groups mentioned above but even from the influence in form of positive or negative attitudes in the family, too hard upbringing and mobbing (Lidman Magnusson, 1994).
Own previous studies

My earliest investigations about "Dysfunction in the Singing-ability-System " in grown-ups (Lidman Magnusson, 1988) have primarily tried to show "profiles" of people with DSS. Seven persons with DSS had exercise privately ten times. The investigation showed that they were able to learn more about singing and develop their singing ability. A further investigation about the appearance of DSS and its background was presented in Lidman Magnusson (1994). The purpose with this study was to describe possible causes and relations between factors who can contribute to DSS. Ten persons were chosen out of 83 persons answering a questionnaire and invited to an interview and an evaluation of there singing voices. Before the interviews the participants sang a song chosen by themselves and an obligatory song ("Ba, Ba Black Sheep"). The latter song was also performed after the interview. Their singing was evaluated by five judges (professional musicians and amateurs) using five scales:

* Total Impression
* Ability to Sing in Tune
* Potential for Vocal Dynamics
* Rhythmic stability
* Voice quality

Regression analysis showed that the Total Impression was best predicted by the Ability to Sing in Tune. These previous studies suggested several psychological and social factors that can interact and make hindrances for early singing development. Therefore a new investigation should investigate the relevance of such factors as circumstances during the individual's growing up period, personal abilities and characteristics, positive reinforcement etc and further try to make comparisons with people who are singing and playing and thus not show DSS. Are there differences and in which way? The purpose of the new investigation was principally to describe similarities and/or dissimilarities between different groups of people concerning circumstances while growing up, individual abilities and characteristics, positive and/or negative influences and other factors of importance for singing development and further to examine the relations found in earlier investigations using a considerably larger number of people.

METHODS

Questionnaire

A comprehensive questionnaire was constructed based upon the results from the two other investigations (Lidman Magnusson, 1988, 1994). It comprised 120 statements, divided into 13 different categories. The categories refer to Circumstances during the childhood, Music activities in the family, Music listening habits, Singing habits, Music in kindergarten and in compulsory school, Negative judgments concerning one's singing, Leisure time activities, Music activities as grown-up, Social relationships, Positive/negative influences, and Statements concerning the own personality. The participant indicated his/her agreement with the respective statement using a five steps scale ranging from Very good agreement (5) to No agreement at all (1).

Participants

The purpose was to distribute the questionnaire to groups of different age and occupation with both genders about equally represented. This may make it possible to compare people who can be described as having DSS with people who have not (such as professional musicians and singers), and compare people who do not perform music with professional musicians. Usually the participants completed the questionnaire privately in their homes; students made it during class. Time required was about 20-30 minutes.
RESULTS

About 500 questionnaires were distributed, and 425 (85 %) were answered. Out of these 318 (75 %) were female and 107 (25 %) males, age range 18 - 70 years. The statistical treatment comprised usual descriptive statistics, further correlations, factor analyses and regression analyses. The different statements in the questionnaire were grouped into two super-ordinate categories, referring to possible causes of and to the possible effects or indicators of DDS, respectively; this reflects the original construction of the questionnaire.

There were five groups of statements indicating possible causes of DDS: called SOCIAL RELATIONSHIP PSYCHO-SOCIAL UPBRINGING, MUSICAL SOCIALIZATION, INFLUENCE (positive/negative influence from family, school etc), and SPEECH AND HEARING.

Four groups of statements referred to effects: "SELF IMAGE", RELATIONSHIP TO MUSIC, EXPERIENCE OF ONE’S OWN VOICE, and SINGING BEHAVIOUR.

The statements in each of these nine groups were subjected to factor analysis (component analysis followed by varimax rotation) to reduce them to a smaller number. For instance, the group INFLUENCE originally consisted of nine statements referring to schoolrelated positive influences, positive music influences by parents, general positive influences by parents, criticism of one’s singing and several others; after factor analysis these could be reduced to four variables.

Some examples of correlations within the respective groups in the "cause" category are as follows. The largest correlation in the group INFLUENCE is between Schoolrelated Positive Influence and Critized (r = -.472); in PSYCHOSOCIAL UPBRINGING between Shy-Melancholy and Afraid (.422); in MUSICAL SOCIALISATION between Own Singing and Mother’s Music Activities (.407) and between Own Singing and Father’s Music Activities (.366); the correlation between Own Singing and Music in the Family is still higher, .588); in SOCIAL RELATIONSHIPS there are high correlations between Relationship to father and Relationship to mother (.446) as well as between Relationship to mother and Relationship to family (.592), and between Relationship to father and Relationship to family (.605). No significant correlations were found between variables in TALKING AND HEARING.

In the "effects" category the highest correlations within groups are the following. In "SELF IMAGE" the highest correlation is between Positive self image and Sensitive to Criticism (-.440) and, conversely, between Negative self image and Sensitive to Criticism (.481); in RELATIONSHIP TO MUSIC between Listening and Liking music and Musical Consciousness (.306); in EXPERIENCE OF ONE’S OWN SINGING between Singing feels good and Hoarseness/Voice fatigue (.341) and between Bad Singing and Hoarseness/Voice fatigue (.358); in SINGING BEHAVIOUR there is a high negative correlation between Singing and humming and Stopped singing (-.554). On-going analyses of the relations between "cause" and "effect" groups include regression analyses and LISREL-type analyses will be reported elsewhere.

DISCUSSION

Singing involves a lot of physical and mental processes, working together in a suitable way, so that you can talk about an ability of "singing". The ability of singing, as well as singing behaviour thus depend on a dynamic system of "working together". If the development of this system is inhibited in such a way that the dynamics in the system is disturbed, it does not help even if the fundamental mechanisms exist. There are different types of interactions between the individual and the environment. They tend to develop or to prevent the development of singing ability by facilitating or blockading the necessary dynamics in the system of "cooperating factors" of significance for singing behaviour. Each kind of DDS may be related to something in the personality and in a dimension which can influence the individual’s experience of oneself. If we can find a cure for this situation this can bring rehabilitation for much more than "only" singing ability.
Earlier investigations about singing ability have often been focused on disturbances of one or another of these functions and even, but to less extent, on motivational aspects. The present investigation attempts to find out how the dynamics in the system of interactions that develop an active singing ability look like. A tentative model of the dynamic relations between factors influencing singing development is shown in Figure 1.

**Figure 1**

*The Dynamic Relations between Factors Influencing Singing Development*
REFERENCES


Torell, Hj. (1945) Om behandling av brummare.[About how to do with droners ] Skolmusik 1945:

Opera Singers: Voice, Psyche and Soma

Maria Sandgren
Department of Psychology, Uppsala University, Sweden

The demands on opera singers’ performance are high, technically as well as aesthetically. Elements of music, singing and acting are involved in the performance. By means of the voice and the body the artists vision is transferred to the audience. Having the instrument within the body may complicate the relation to the body and the psyche. It may as well have consequences for the private life.

The instrument - the voice - is of utmost importance for the singers for many reasons. It is their very individual expression of artistry and the core of the profession. Singing requires a good physical health and any malady may be causing inconvenience. The main worry for the singers mostly concerns somatic problems which affect the vocal ability (Sataloff, 1991). The attention directed inwards as an effect of having the instrument in the body is a characteristic of the profession. Other musicians may project their problems on to the instrument, but the singers have to take charge themselves of defects as well as assets. A certain body sensitivity develops in order to make the invisible instrument in some way detectable and visible (Kemp, 1996).

Psychological problems may as well be reflected in the voice such as depression, insecurity and other emotional disturbances (Sataloff, 1991). Artists were found to have high scores on emotionality: dependence, hypocondria, guilt, anxiety which predicted depression, feelings of panic and migraine (Marchant-Haycox and Wilson, 1994). Vulnerability in self-esteem could be an important factor. If the need for approval and confirmation is too dominant, the singers will be vulnerable for feelings of emptiness, shame, ugliness, inferiority as the self-esteem is depending on validation from others outside. Feelings of strength and success would be of considerable importance in order to maintain self-esteem (Kernberg, 1975; McWilliams, 1994). High demands on perfectionism (setting unrealistic standards for oneself as well as significant others) may develop which can be associated with somatic complaints (Organista and Miranda, 1991), depression (Hewitt and Flett, 1991) and performance anxiety (Montello, 1992).

The impact of psycho-social factors may also be great. Artists (musicians, dancers, actors) are exposed to many sources of stress as irregular working hours, travels, job insecurity, difficulties combining professional life with family life, progress impatience (Ramel and Moritz, 1991; Schmale and Schmidtke, 1985; Wills and Cooper, 1988). The way the life of a successful opera singer is made public may become a strain. The opera singer Jussi Björling (1994) found that acknowledgement and success were followed by gossip, pedantry and often unfair criticism. He questions if the way his life was made public - "every little detail was examined" (p. 16) - was the price to pay for success.

There are several factors that may inhibit the opera singers in their strivings for artistic and technical perfection. As research about opera singers is very scarce, the aim of this study was to investigate problems and coping strategies used in the professional life of opera singers. Moreover, the singers were asked about positive singing experiences and motivational factors. In the second part of the study, specific problems (vulnerability in self-esteem, depression,
preoccupation with the voice etc.) were investigated.

**Method**

In the first part of the study, semi-structured interviews were conducted with 15 opera singers, 5 males, 10 females, age 27-69. The singers were asked to tell about problems and coping strategies used in their professional life. Furthermore, motivational factors and positive singing experiences were studied. Semi-structured interviews with 15 persons who were in regular contact with opera singers (conductor, singing teacher, physician etc) were also studied in order to get a wider perspective on the professional life of opera singers. Moreover, the author visited an opera-house backstage during a performance for gaining more insight in the process of the performance.

The results from the interviews brought forward the construction of a questionnaire consisting of 138 items and two open questions. The questions concerned demographic data (sex, age, civil status, education, terms of employment), vocal pitch, vocal training, contact with caregivers, somatic problems, depression, activities to promote health, hypochondriac tendencies, preoccupation with the voice, vulnerability in self-esteem and performance anxiety. Three variations of a five-point Likert-scale was used where the subjects could indicate their agreement with various statements. The intended subjects were 72 professional opera singers divided in four strata: males and females equally distributed in the two groups: freelancing and permanently employed. The final sample included 14 males and 13 females freelancing as well as 10 males and 13 females permanently employed, age range 21-60 years. Response rate was 67 percent.

In order to study sex differences as well as differences between older and younger, freelancing and permanently employed singers t-tests were conducted. Correlations were used to study relations between depression, vulnerability in self-esteem, performance anxiety, preoccupation with the voice and somatic problems. A factor analysis was performed regarding promoting activities.

**Results**

The results from the interviews gave evidence of a considerable amount of problems - somatic, psychological, psycho-social - of which only a smaller selection will be presented. As expected, somatic complaints concerned above all problems in the vocal organs and respiratory tract which could cause indisposition. Reoccurring periods with somatization (i.e. psychological factors, maladaptive behavior etc. adversely affect the general medical condition) were frequent. In order to stay in good physical condition, the singers used health food and avoided environments/situations containing harmful stimuli for the voice and/or the body. They stressed the importance of developing a good vocal technique and the use of regular vocal training which would prevent them from developing vocal difficulties.

Psychological problems regarded in first hand hypochondriac tendencies and vulnerability in self-esteem. Preoccupation with the existence and quality of the voice seemed to be more prevalent for males than women. The way of testing the voice was associated with maintaining self-esteem. If the voice was functioning well, the singer felt assure and competent. On the other hand, when it failed, it often caused feelings of shame and inferiority. Moreover, the singers had ambivalent feeling towards both positive and negative criticism. It appeared as if receiving criticism was as one of the most considerable strains in their
professional life. This ambivalence can be seen as a strong need for receiving confirmation and approval.

In order to stay in mental balance the singers emphasized the need for relaxation and recreation. Some used mental training and meditation. They appreciated activities without any connection to opera, for example walks and family life. The singers pointed out the importance of having a good vocal technique in order to, for example, control the effects of performance anxiety.

Regarding psycho-social problems, a majority of the singers stressed the difficulties of separating professional life from private life. The professional life easily became too important and the singer "lived his/her life on the stage" instead of "working on the stage." Therefore, the importance of maintaining a family life was emphasized. Irregular working hours and freelancing made it hard to combine family life and professional life as well as had consequences for their partners' careers.

There are, of course, also bright sides in the professional life of the opera singers. The audiences' response to the performance was of crucial importance. The singers wished that the audience would accept and understand their artistic visions and talents. If this succeeded, they felt satisfaction with feeling confirmed and understood. Experiences of mastery and skill on stage were frequently mentioned as highly motivational factors. The males expressed great satisfaction over moments when they had felt completely free to do whatever they wanted with their voice, technically as well as aesthetically. A need to express emotions was mentioned mainly by the women. Strong emotional music experiences with transcendental elements were also found.

The results from the questionnaire indicated that vulnerability in self-esteem was significantly and positively correlated with a number of variables such as somatic problems, performance anxiety, preoccupation with the voice. The correlation between depression and vulnerability in self-esteem was high. Muscle tensions and fatigue of the voice were correlated with need for approval and confirmation as well as feelings of insufficiency. Other somatic problems was correlated with tiredness, weariness, inability to perform according to one's capacities, doubts about one's artistic capacity and the negative impact of work on private life. The latter indicates signs of depression. Too high consumption of food, sweets etc. was also correlated with vulnerability in self-esteem and depression.

Performance anxiety was highly correlated with vulnerability in self-esteem and depression. In contrast to earlier findings about females experiencing more performance anxiety than males (Wesner, Noyes and Davis, 1990), very few differences were found in this study. However, more intriguing were the positive correlations between performance anxiety and depression as well as vulnerability in self-esteem which were stronger for males than for females.

The singers tended to test the existence and quality of the voice several times a day. As vocal testing was correlated with health concerns and vulnerability in self-esteem, this can be regarded as hypocondriac tendencies as well as self-regulation of self-esteem. It seemed as females tested the voice because of fear of illness, while men's way of testing seemed to be more correlated with self-regulation of self-esteem. Men also tested the voice more often than
the females did.

The singers used various coping strategies in order to maintain physical health and mental balance. Some of the strategies involved irrational beliefs about the effects. No correlations existed between absence of somatic or psychological problems and health promoting strategies.

Discussion
Opera singers experience a wide range of problems in the pursuit of a professional career. The way of using the body and the voice as instrument has several complications. The voice plays an important role as it is said to reflect the singer's potential and defects. Thus, the singer's mental balance often depends on the quality of the voice. The singers' difficulties to separate private from professional life is of great concern. The need for approval and confirmation has a strong impact, irrespectively of somatic, psychological or psycho-social problems.

The psychological meaning of using the body as medium of the art has implications. The singers may feel that the body is exposed to various threats such as illness, others judgements etc. It appeared as if they had an individual repertoire of harmful stimuli as well as individual coping strategies of more or less irrational kind. However, the complete exposure in front of an audience involves both psyche and soma. How the singers dealt with the constant evaluation of the performance by the audience, significant others and, not at least, by the opera singers themselves had a strong impact on self-reliance and self-esteem. Failure often meant loss of self-esteem which equated failures and mistakes with being a failure as a person. This could lead to a demand for perfection in order to maintain self-esteem. Perfection should not only be regarded as harmful, but also a way of setting personal standards for oneself. The latter is associated with positive achievement strivings and work habits. In that case the singers would not be driven by fear of failure but work out of enjoyment and deep personal involvement.

The results from self-reports can be questioned. In this study, none of the singers reported serious difficulties with performance anxiety when they were asked to indicate their experience of it before and during the performance. One possible reason to the low scores may be that the singers denied the severity of performance anxiety. However, the scores of physiological and cognitive reactions indicated greater severity, but in no case extremely high. Considering somatic problems, there was a great variation but of low severity. Problems with alcohol and drugs were non-existant. There is reason to believe that singers who are suffering from severe problems did not answer the questionnaire.

How these problems affect the performance and the artistic development are not answered in this study. In the next few years a longitudinal study involving pupils at an opera school will offer the possibility to more exactly pinpoint crucial interactions between environmental and personality factors in order to strengthen the artistic identity. Furthermore, it will be an opportunity to examine important developmental stages during the artistic education.

References
Music Analysis: Composers, Works, and Genres
Emotions, Attitudes, Images and Ideas in J.S. Bach’s Music of the Imagination

ELISABETH KOTZAKIDOU PACE
Columbia University

In this study I provide an analysis of J. S. Bach’s Coro I.II. and duet “So ist mein Jesus nun gefangen” from his St. Matthew Passion. This exercise turns out to be interesting in two respects: Firstly, what is revealed is the extensive and purposeful employment of musical keys as rhetorical figures, which the composer employs to express the emotional content of the text in a way that highlights its theological intricacies. This is hardly ever reported in the relevant literature (Kivy, 1984 and Schweitzer, 1911 among others). The existence of different key characteristics (let alone their subjectively discerned sensory character) is, of course, completely lost to the modern listener, who in most cases has to contend with an equal-temperament performance. Secondly, the very form of the remaining rhetorical figures (which range from the particular choices of time signatures, melodic shape, compositional procedures, sound textures, pitch classes, and rhythmic patterns, to performance articulations) is significant, because, when examined in light of recent developments in cognitive semantics, they seem to be meaningful beyond what the more traditional semiotic analysis would have us believe. The figures in question have indeed a semiotic function. Yet the actual symbols themselves, conventional though they may be, and in wide use during the baroque era, are not arbitrary. They embody a “logical form” that seems to correspond to the logical form of the meaning of the words they are “illustrating”. For that reason we can properly call them “cognitive schemata” and take an interest in the particular ways in which they seem to form connections with other cognitive domains (or “mental spaces” in Fauconnier’s 1985 sense) from which they “import” the structure (or “logical form” in Susanne Langer’s, 1957 sense) of their meaning.

An Analysis: J.S. BACH St. Matthew Passion; Soprano & Alto Duet with Coro I.II. “So ist mein Jesus nun gefangen”

Duet (4/4):
So ist mein Jesus nun gefangen
Mond und Licht ist von Schmerzen untergangen
weil mein Jesus ist gefangen
Sie führen ihn, er ist gebunden

Thus is my Jesus captured
moon and light have from pain succumbed
because my Jesus is captured.
They lead Him away, he is bound

Choral Interection:
Lasst ihn, haltet, bindet nicht!
Let him go, stop, don’t bind Him!

Coro I.II. (Vivace 3/8):
Sind Blitze, sind Donner in Wolken verschwunden?
Blitze, Donner, Blitze, Donner...?
Eröffne den feurigen Abgrund; o Hölle
zertrümme, verschlinge, mit plötzlicher Wuth
den falschen Verräther, das mörderische Blut.

Has lightning, has thunder in the clouds disappeared?
lightning, thunder, lightning, thunder...
Open up the fiery abyss, O Hell
shatter, devour with sudden rage
the treacherous betrayer, the murderous race.

Duet. As an immediate reaction to Jesus’ betrayal and arrest comes the duet “So ist mein Jesus nun gefangen” with choral interjections which eventually erupt into a double chorus piece “Sind Blitze und Donner in Wolken verschwunden?”. The duet portion is scored for (Flute I & Oboe I) and (Flute II & Oboe II) anticipating the texture of the (soprano & alto) duet, with violin I, II and viola playing in unison. There is no basso continuo, the absence of which underscores the feeling of loneliness and emptiness as the Savior is snatched away. The continuo does not appear as an independent line until the choral portion is introduced. The key of the beginning is e minor, which

Copyright 1997 by the author.

351
is described in historical treatises as the key of “tender and innocent declaration of love, of lament without grumbling” (Weikert, 1827) (1), and is thus perfectly suitable for the expression of resignation and acceptance of a painful fate. The two voices start out in a canon arrangement suggesting that Jesus is leaving. This caccia texture was a widely used conventional symbol for leaving, going, chasing or leading. Since this texture is also employed later for a more pictorial description of Jesus’ literal “being led away” scene, this first appearance of a canon seems to have more of a psychological, emotional significance, and it is meant to evoke a sense of separation and being left behind, upon Jesus’ imminent departure. The melody has a descending line suggesting resignation and a slow tear rolling down the cheek. In addition, the downward movement away from the tonic while avoiding any leading tones (8–6–5–4–3) creates an unsettling feeling of separation anxiety. The slow dotted rhythm in 4/4 meter is traditionally considered one of dignity, solemnity, and quiet despair. For the whole canon portion, but especially on m22-23 on the word "gefangen" (captured), the alto voice joins the soprano giving rise to a combined rhythm which strongly suggests entanglement and the twining of a rope around the hands of Jesus as he is being arrested.

At this point a b minor 6-3-5 chord is violently injected by the chorus of followers in abrupt stac­cato articulation: “Lasst ihn, haltet, bindet nicht!” (Let Him go, stop, don’t bind Him). The key of b minor - notice that Bach didn’t use a simple V chord which would have been B major - is considered to be a “biting, dry, and savage” key (Castil-Blaze, 1821), in which “black and gloomy” (Knecht, 1792), “witch-like” (Anon, 1828) sentiments can be expressed. Notice then the juxtaposition of the dignified, quiet resignation of the e minor feeling of the duet with the savage b minor shouts of the crowd of Christ’s “followers”: the duet is expressing the proper pacifist, resigned, quiet acceptance of God’s will (which is appropriate to a highly indoctrinated Christian), while the chorus sounds like the deeply emotional and very threatening screams of a mob demanding instant and vigilant justice. Since their indoctrination into Christianity has not yet been completed, their natural heart demands justice for Christ through violent means; after all Christ has not yet gone to the cross. The choral vivace that follows the duet is an intensified form of this wild rebellion of the “subconscious” crowd: a shouting out of curses that old village women still do around the Mediterranean in order to decry injustice.

Meanwhile, the two women present themselves in a first person narration (“mein Jesus”), which places them in the foreground, the center-stage position of the mental space created by the text. Next, notice the off-center-stage point of view of the outraged chorus of Christian “sympathizers” as it is shouting first demands for the release of Jesus and then curses against his arresters and their whole race. The arresting Jews themselves are completely backgrounded, off-stage. What is the issue here is not “their” actions, but the way “we” Christians are supposed to react. They are made irrelevant, since the true conflict is not between the Christians and the Jews, but between the true Christians and those who think they act on behalf of Christ. What emerges then as theologically significant are not all three personae participating in this religious drama, but the two moral points of view espoused by the foregrounded and the middlegrounded players. These two opposing moral forces are the totally internal to the faith, personal “mine” point of view of the two women, and the off-center “us”, with its morally inappropriate “us versus them” attitude. The audience is ultimately expected to “do the right thing” and pick the two women to identify with, but not before they are offered, in the Vivace, the psychologically sound opportunity to “purge” their anger through the safety valve offered by the slightly removed, off-center persona of “them”, the “primitive”, “overzealous”, “wet-behind-the-ears” chorus.

In m27 the duet returns homophonically in thirds to show unity in Christ. Descending “sighs” (eighth notes in groups of two with connected articulation) musically illustrate the “grief”-word “Schmerzen”. The whole sentence reads “Moon and Light have from grief disappeared”, so that the text’s main thrust is the disappearance of light, and ultimately the upsetting of the natural order

(1) The quotes of historical treatises used in this analysis were taken from Steblin’s 1981 key characteristics index.
of things. Bach chose to set the passage in B Major and end it in E major. We are told poignantly that B Major is “a harsh key, more piercing than E, fit to express cries of despair, howls, roars and the like” (Galeazzi, 1796) and that “the state of nature has already disappeared to some extent” (Heinse, 1795). Both B and E Major are “the most glaring coloring; they are to be compared to shining fire-colour and burning yellow” (Seidel, 1828). This last description foreshadows the use of E Major later in the piece. For the time being, in combination with B Major, it is a description of the light that went under, and the ensuing upset of the natural order.

On m36, the “entanglement of the rope” motive on the word “gefangen” (captured) returns. On m45 Bach starts the “leading-away” figure, upon the words “sie führen ihm” with the alto leading the canon. The ascending line, which even trips downwards momentarily, with its repeated notes and the dual articulation sounds like Jesus’ dragging and stumbling uphill steps as He is pushed and pulled by the soldiers. On the word “gebunden” (tied up), where both the intrinsic meaning of the verb and the past perfect tense suggests completion of an action and achievement of a steady state, a long single note is held steady over two and a half measures. Starting with m50, on the word “führen” the step-wise descent of the vocal line over the unsettling b minor pedal of the strings in unison comments on this morose and gloomy moment. Bach then returns to the canon treatment of “führen” which on m65 literally leads us into the b minor double chorus.

Coro I.II. In sharp contrast to the previous 4/4 measure of dignity, we are now in a fast paced 3/8 time signature, which in combination with the first entrance of the basso continuo in a motive of six sixteenth notes creates great agitation. This further combined with the two choruses singing first homophonically, then antiphonically, creates a wild acoustical image. The words “thunder”, “lightning” spat out in a crossed pattern by the two choruses sound percussive, as if real thunder and lightning are striking against a constant stormy background, where even the clouds are depicted by the thick texture of the strings (on the printed page, if not in the domain of sound as well). The whole scene keeps building up, and as the texture gets thicker and more belligerent, the mounting storm becomes a metaphor for the inner rage and frenzy of the heart. Since the two choruses become antiphonal on m95, it may be reasonable to assume that a spatial separation of the two groups in performance is desirable, in order to create the additional effect of being surrounded by the raging storm: the “Blitze” and “Donner” are hitting us from all directions until the fermata of m104. It is then that the gates of Hell are expected to open with a witch-like call in F sharp major: now the percussive sound of the thunder and lightning acquires the double meaning of crazed knocks on the gates of Hell. The key of F sharp major is said to be “very brilliant, piercing” (Reicha, 1814), “sharper than its enharmonic G flat major and yet expressing dark feelings” (Ebbhardt, 1830), while the fact of its enharmonicity to G flat major makes its usage an “ambiguous vacillation between heaven and earth” (Mueller, 1830)! In addition, “where the most inner feeling has worked its way out of pain and violently seizes the joy which has eluded it, this in our opinion, is F sharp major” (Schilling, 1835). What more appropriate key than this to express this unrestrained outburst of the subconscious “Id” violently surfacing and taking over the restrained resignation of the preceding duet (“Ego”). “Id” is calling for the gates of Hell to open up and swallow the perpetrators of the injustice: “may you burn in Hell” is what is told in effect. On the percussive words of the curse “verderbe” (may you rot), “zerschelle” (shatter), “zertrümmre” (be crushed), “verschlinge” (be devoured) starting on m121, notice that the six sixteenth motive of the bass has evolved into a terrorizing figure of repeated sixteenth notes (possibly alluding to Monteverdi’s “guerra” figure), which chromatically descend to Hell. On m130-31 the intoning of the words “falschen” (treacherous / false / liar) and “Verräter” (betrayer) are particularly interesting. The “speaking out of both sides of the mouth” quality of “falschen” is depicted by an ambiguous diminished seventh chord (which is conventionally considered rootless and thus “without moral foundation”), while the attitude of the Chorus towards the “Verräter” is made plain with a B Major chord, which is said to “have an offensive, hard, unpleasant and also somewhat disparate character” (Mattheson, 1713). On the words “das mördische Blut” (murderous blood) an attitude of disdain and disgust is evoked by the e minor chord, which in this context is described as “pale orange” (Ribock, 1783) and “almost banished from music of good taste, save for modulations”
This terror scene, which had already given us short glimpses of fire in m113 and m117 with an E Major on the word “feurigen” (fiery) and a piercing bright A Major on the word “Hölle” (Hell), now ends up in an E major cadence which, far from being an innocent Picardy third fulfilling a stylistic requirement, is the key that “can depict fire best of all, especially as it stands out through the intensity of its piercing flames” (Vogler, 1779). This leaves no doubt that the gates of Hell have finally opened up!

Overall, the juxtaposition of the duet with the double chorus speaks in no uncertain terms of the dawn of the New Testament’s “turn the other cheek” (which has not yet taken hold in every new convert’s heart), while the Old Testament’s “an eye for an eye and a tooth for a tooth” is waging its last (losing) battle to assert its validity as the corner stone of the divine Law of Justice.

Some Cognitive Schemata found in the musical example.

a. Key characteristics. The current consensus of expert opinion is that the actual perception of differences between the keys in the Baroque era should be attributed to the (various) unequal well-tempered tunings employed by J.S. Bach and his predecessors in Germany and France. Unlike today’s “equal-temperament”, a well-tempered unequal tuning would retain an extra dimension of richness-in-variety that the ideal of writing in all twenty four keys sought to provide. A host of historical musical authors listed by Steblin, 1981 provide detailed information regarding the typical “intrinsic” key characteristics with a remarkably high statistical level of consistency. The specific source of the key characteristics (acoustical properties of the tuning system, poetic associations stemming from actual usage, or a combination of both) is immaterial. What is important is the fact that these associations were highly conventional and were used freely as a vocabulary. In fact, Bach’s felicity of key employment in diverse contexts lends independent support to this notion.

The music of the High Baroque is the kind of music that was purposefully written to be a “language”, the latter notion to be construed in accordance to the linguistic theory of the time: Rhetoric. J. S. Bach elevated this inherited art form to a truly personal art. Through innovative usage of preexisting conventional signs and the imaginative creation of novel ones, he expressed the most complex intricacies of both human psychology and religious doctrine. As our musical data shows, the choice of key illustrates the moral character of the personage or the attitudes of one drama player towards another, evokes colored images of fire and blood, and semiotically points to even highly abstract metaphorical links, such as “the natural order of things” or “vacillating between Heaven and Hell”. Even the otherwise trite employment of the Picardy third (which turns the last chord of the piece into E Major) acquires a novel meaning in the hands of the master: he actually supplements the text with an image that conveys the stereotypical depiction of Hell as a place of fire. First we get a passing glimpse of a few flames piercing through until, finally, we are suddenly exposed to their full terror, as the gates of Hell open wide.

b. Compositional procedures. Let us take the example of fugal writing which conventionally is used to represent the related ideas of “leaving”, “following”, “leading” (“leads into”) and “leading” (“they are leading Him away”). The way Langacker’s, 1987/91 Cognitive Grammar would describe the basic meaning of this family of verbs would be to evoke a trajectory, namely a directional path, along which two entities (say, the subject “they” and the object “Him”) are represented engaging in a certain action of one upon the other. The whole scene involves a point of view, with respect to which the trajectory of motion assumes a direction (in this case “away from”). A whole family of verbs can stem depending on which aspect of the basic scene is foregrounded in the schematic representation of the action. It is now easy to see why the texture of a two voice canon would be an appropriate symbol pointing to the salient features of the schematic representation of the meaning of the verb-family of “leaving”. The two mental spaces (Fauconnier, 1985) are in partial correspondence, which allows for this metonymic relation to take place. What is intriguing and theoretically significant for Cognitive Science is the fact that the two connected
mental constructions are manifested in different media: one represents the “logical structure” of a linguistics text, and the other the “logical structure” of a musical “text”. Similarly, the homophonic writing employed to evoke a sense of “unity” and “togetherness” with Christ succeeds in its communicative goal through the obvious mental link that is easily established between the space containing the musical material of parallel thirds and the salient features of a cognitive schema of “togetherness” (stereotypically depicted as two people in synchronized motion).

c. Rhythmic patterns. The musical realization of the word “gefangen” (captured) involves a more elaborate connection between the musical and the linguistic schemata. The “capturing” scenario (in Schank’s, 1975 sense) involves many options regarding the choice of instrument. A modern filler of the instrument-gap in the scenario would be a set of handcuffs. In that case the entangled combined rhythm created by the two voices surely would have been an infelicitous musical representation of that concept. Instead, Bach supplements the text (which does not actually specify the instrument) by filling in the stereotypical instrument of his era: a rope wrapped around Jesus’ hands, so that the twists and turns of each of the melodic lines and their rhythmic “entanglement” do actually capture what is salient in the schema of a twining rope. Similarly, the use of a long steady note on the past-participle “gebunden” to depict the achievement of a steady state is equally intriguing, considering that Langacker analysis’ shows past-particples to be constructions which foreground the steady state achieved at the end-goal of the verb schema “start-path-goal”.

d. Melodic shape. The use of ascending or descending melodic lines (in combination with appropriate performance articulations) to express the upward dragging footsteps of Jesus, sighs, tears rolling down the cheeks, the descent into Hell, or the undulating dance of the flames are conventional signs in wide use throughout the Baroque era, whose success as signs depends on the obvious correspondence of their structure to the structure of the kinesthetic images they are evoking. Mark Johnson, 1987 has argued for the role “prelinguistic” kinesthetic images play in the creation of linguistic meaning. Bach’s musical constructions provide us with marvelous examples of the same cognitive phenomenon in the domain of music.

e. Tonal material. An interesting case of a musically “internal” sign is the employment of a pitch class containing no leading tones, both to represent the unsettling feeling of not knowing which way to turn (now that Jesus, our guide, is taken away) and to evoke an actual feeling of anxiety in the hearts of the audience. The success of such a ploy depends on the prior expectation of the listener that clues should be present which can be trusted to lead her back to the restfulness of the tonic, in addition to the expectation that any movement should be towards the tonic - not away from it. The very fact that our language forces us to describe this musical situation employing metaphors such as “lead him to the tonic”, “arriving at the tonic”, “the tonic is a point of rest”, “harmonic movement away from it creates a yearning for the tonic” is significant. Not only are we conditioned by the structure of the tonal system itself, as it appears in the domain of sound, but our linguistic structuring of it, the way we talk about it, evokes the idea that we are left without a guide, should the leading tones be missing. It appears that a multi-level interaction takes place between the mental space of sound representation and the mental space of our linguistic representations, that gives rise to the ingenious result Bach intended. Similarly, extended chromatic movement can be very unsettling and anxiety provoking.

f. Tempo. Tempo specifications combined with certain time signatures (such as the figure of slow dotted rhythms in 4/4, which is said to be dignified, or a fast 3/8 rhythm, which is said to be agitated) also seem to have a kinesthetic foundation. Whether the association gives rise to polyseme is irrelevant, since this latter phenomenon is pervasive in natural language anyway, and in most cases is resolved by the context.

g. Texture. The texture of the orchestration can also be very expressive. Thus, the thickness of the strings is employed to create the image of thick clouds. This is an example of what Kivy, 1984 calls a “notational” sign, namely something that is meaningful on paper, but may or may not
translate well into sound. Kivy offers the humorous example of a pure notational sign Telemann employed when he wrote his Gulliver Suite: the Lilliputsche Chaconne is written in tiny note values, while the Brodindnagische Gigue is written in enormous note values. Furthermore, the antiphonal employment of the double chorus creating percussive sounds is an instance of Kivy's "sounds-like" category of musical meaning creation. The listener feels as if she is plunged into the midst of a raging storm. But the most intriguing example (not included in Kivy's categories) is the case where the absence of the conventionally expected basso continuo is set in correspondence with a mental space where the absence of a person is highlighted. As psychoanalysts from Freud to Lacan know, our mind often takes advantage of even more obscure relations of homophony, polysemy and broad metaphoric relations between words, to draw interesting and unexpected associations. For example, in a dream unfolding on the "banks" of a river, the dreamer may be secretly alluding to an episode in her life that took place at the financial institution of a "bank".

h. Complexity. According to Langacker, an important aspect of meaning formation is the ability of schemas to interact and combine to form complex schemata. The way cognitive schemata from all levels of the semantic construct come together to delineate the characters, their point of view, and their status as center-stage, off-center-stage, or off-stage is indicative of their immense potential for "building up" intricate relations within a musical mental space. Another example of compositional complexity is the synergism of the "leader-less" melodic line (evoking the feeling of "separation anxiety" and lack of direction) with its descending trajectory (expressing the sadness of the speaker), while the key of e minor (the key of "tender and innocent declaration of lament without grumbling") and the meter of 4/4 (the signature of dignity) further reinforce the meaning of the scene. Also worthy of mention among the figures we examined is the situation where, through some musical device, the audience is urged to assume an attitude towards the scene or one of its characters. For example, the intonation of the word "betray" using an "ambiguous/rootless" diminished seventh chord, or the way the audience is subtly urged to identify with the two women, through the serenity of the e minor key.

i. Context dependency and metaphorical extensions. In closing, notice that each key seems to be open to a positive as well as a negative way of interpreting its characteristics, and each schema seems to acquire new meaning depending on the context. A descending melodic line can mean running tears or the descent into Hell, while the key of e minor, which in its positive version is the key of quiet resignation and acceptance without grumbling, seems to also have the negative reading of "pale orange color" (of spilled blood), and is thus "banished from music of good taste". This phenomenon alludes to the metaphoric and context dependent nature of musical meaning, which cognitive linguistics has also shown to be the essence of our linguistic function. Just like the schemata evoked by scientific texts, the spatial prepositions of a natural language, or the spatial schemata evoked when we are trying to give directions to somebody's house, these musical schemata seem to be just another expression of our cognitive "semantic" function caught in the act of employing the same useful strategies of meaning creation in a variety of domains.

REFERENCES

De Rome, qu'il s'apprétrait à quitter après avoir passé deux années à la Villa Médicis, Debussy écrivait le 9 février 1887 à Emile Baron, libraire parisien qui le tenait au courant de la vie littéraire et artistique dans la capitale : "Ah ! Envoyez-moi "La Nouvelle Revue" du 15 février qui doit contenir un roman de Bourget". Bien qu'elle fût inexacte, car il s'agissait en fait de la publication de six sonnets, cette demande n'en était pas moins révélatrice de l'intérêt que le jeune musicien portait alors à son aîné de dix ans qui, après avoir publié en mai 1882 son dernier recueil de poésies, "Les Aveux" et suscité un débat considérable autour de la parution en octobre 1883 de cinq articles de critique littéraire regroupés sous le titre de "Essais de psychologie contemporaine", allait rapidement accéder à la notoriété en 1889 avec son cinquième roman, "Le Disciple". L'étude chronologique des neuf mélodies que Debussy composa à partir de "Les Aveux" semble épouser cette évolution ; en effet, aux six premières, datées de septembre 1883 à février 1884 et dédiées Madame Vasnier, succèdent les trois dernières, vraisemblablement ébauchées lors du séjour romain et publiées en 1891 chez deux éditeurs parisiens : "Beau soir" chez Vve Girod, et "Deux Romances", chez Durand.

Nous commencerons par nous interroger sur le choix littéraire auquel procède Debussy ; une évidence s'impose à la simple énumération des titres choisis : six d'entre eux sont une allusion directe soit à la "Musique" elle-même (Livre I, p. 50), soit au genre de la "Romance" (Livre I, p. 16 et 51 ; Livre II, p. 71, 79 et 101). De plus, si l'on regarde la trentaine de titres que comporte chacun des trois Livres du recueil "Les Aveux", on constate que ce même titre de "Romance" est utilisé sept fois dans le Livre I, cinq fois dans le Livre II, et une seule fois dans le Livre III, ce qui explique très probablement la désaffection du musicien pour cette dernière partie. Cependant, ce terme générique est bien loin de recouvrir une même réalité de sentiments ou de sensations ; ainsi, la plus passionnée de toutes, la première "Romance", datée de septembre 1883, sera-t-elle finalement écartée au profit de
celle plus descriptive datée de janvier 1884, et finalement publiée en 1902 ; mais en 1906, seules les "Deux Romances" de juin 1891, par leurs sensations quintessenciées de sons et de parfums, trouveront grâce aux yeux du compositeur.

L'intérêt soutenu dont ont bénéficié pendant plus de vingt ans ces neuf poésies, aujourd'hui bien oubliées, de Paul Bourget ne pouvait pas ne pas avoir d'incidence sur le traitement musical auquel s'est livré le compositeur ; par-delà les variantes que l'on relève dans "Paysage sentimental" et "Voici que le printemps...", et qui évoquent immanquablement celles plus connues des "Fêtes galantes" de Paul Verlaine, nous tâcherons de mettre en évidence les oppositions tonales et agogiques qui distinguent les six premières mélodies des trois dernières, celles de 1891, pour finalement nous attarder sur l'exceptionnel équilibre qui s'établit entre la voix et le piano dans les deux dernières Romances.

Editées une première fois immédiatement après les "Ariettes oubliées" et les "Cinq poèmes de Baudelaire", ces "Deux Romances" ouvraient la voie aux "Proses lyriques" et aux "Chansons de Bilitès" ; rééditées en 1906 avec "Le jet d'eau", "Fêtes galantes" et "Trois Chansons de France", elles demeurent parmi les réussites les plus rares de Debussy.
<table>
<thead>
<tr>
<th>Chronologie</th>
<th>Manuscrits autographes (Paris, B.N.F.)</th>
<th>Les premières éditions parisiennes</th>
<th>Livre I : Amour</th>
<th>Livre II : Dilettantisme</th>
<th>Livre III : Spleen</th>
</tr>
</thead>
</table>
| Sept. 1883        | Romance : "Silence ineffable..."  
                        | Musique : "La lune se levait"  
                        | Paysage sentimental : "Le ciel d'hiver..."  
                        | Romance, musique pour éventail :  
                        | "Voici que le printemps..."  
                        | Janv. 1884         | La Romance d'Ariel :  
                        | "Au long de ces montagnes douces..."  
                        | Fév. 1884         | Regret : "Devant le ciel d'été..."  
                        | Fév. 1884         |                                                                                                                                  |                                                                                     |                                             |                                             |
| 15 avril 1891     | Paysage sentimental  
                        | Beau soir :  
                        | "Lorsque au soleil couchant..."  
                        | Deux romances :  
                        | - Romance : "L'âme évaporée..."  
                        | - Les cloches : "Les feuilles s'ouvraient..."  
                        | Juin 1891         | Deux romances :  
                        | - "L'âme évaporée..."  
                        | - "Les feuilles s'ouvraient..."  
                        | 1891              |                                                                                                                                  |                                                                                     |                                             |                                             |
| 1902*             | Romance : "Voici que le printemps..."  
                        | Paysage sentimental  
                        | Société Nouvelle d'Editions musicales  
                        | Romance : "L'âme évaporée..."  
                        | Les cloches : "Les feuilles s'ouvraient..."  
                        | 1906**            |                                                                                                                                  |                                                                                     |                                             |                                             |

Notes :  "dans un recueil de "Trois mélodies", respectivement sous les n° 2 et 3  
**"dans un recueil de "Douze chants", respectivement en pages 1 et 21
Schoenberg’s Atonal Pairs

Niall O’Loughlin
Loughborough College of Art and Design
Epinal Way, Loughborough, Leicestershire, LE11 3GE
United Kingdom

The years 1908-13 were difficult ones creatively for the Austrian composer Arnold Schoenberg. His writings make clear how he faced each stage of his artistic difficulties and each musical crisis as it appeared (1). Schoenberg was constantly aware that his move from the received musical idioms of the late 19th century to a new and untried group of compositional methods created a communications crisis between his audience and his music. The principal area of change was in his apparent abandonment of a fixed tonality in the finale of the String Quartet No.2 of 1908. All the works from this time until the first twelve-note compositions in the 1920s (Five Piano Pieces op.23) are generally subsumed under the umbrella term ‘atonality’ (2).

Despite protests by Schoenberg himself against the use of the term, the word has remained as part of the musicologist’s vocabulary. This has the obvious convenience as a recognisable label in identifying a known group of musical works. A simplistic impression of stylistic uniformity, however, has been systematically demolished, especially in the last 30 years, using techniques newly devised or modified to illuminate the creative process and the way that it developed. One consequence has been the ability to make subtle distinctions in this compositional development.

Taking the works that Schoenberg composed in the so-called atonal period (1908-23), but excluding the anachronistic Gurrelieder; the unfinished Symphony and Die Jakobsleiter, the songs op.22, Herzgewächse, and perhaps with less justification the two songs op.14, we are left with four pairs of works composed between 1908 and 1913. A detailed analysis of these works shows the break with the past was not as sudden as has been portrayed and that within this short period a considerable stylistic metamorphosis occurred. If the earlier works is compared with the later ones, an interesting dichotomy results. Of the contrasting pairs in each medium, the first (composed 1908-9) are stylistically ambivalent, ostensibly using an atonal idiom but at the same time showing a backward view by retaining many identifiable vestiges of traditional tonality, motivic usage and formal structures. The pieces in the second group (composed 1910-13) give much less indication of traditional tonality. Their formal structures are much more terse and compressed, with very little or no dependence on traditional plans. The motivic networks tend to be suppressed and modified to the extent that they are usually very difficult to recognise as such. In short, while the first group makes an ambiguous break with the past, the second group of works affirms the composer’s new position with far greater conviction. This appears in works of four categories: for piano, for orchestra, for solo voice and for the stage. For piano there are Drei Klavierstücke op.11 and Sechs kleine Klavierstücke op.19, while for orchestra Fünf Orchesterstücke and the untitled and incomplete orchestral pieces of 1910 (published in 1962 as Three Pieces for Chamber Orchestra). For voice the former is the group of George settings Das Buch der hängenden Gärten and the latter Pierrot Lunaire. The two stage works are Erwartung and Die glückliche Hand. This study aims to elucidate these features. Talking about differing views of the Piano Piece op.11 no.1 Leonard Stein wrote: ‘the ambiguities inherent in this composition are to be found in other works of Schoenberg as well’. (3) Schoenberg himself wrote: ‘Most critics of this new style failed to investigate how far the ‘eternal’ laws of musical aesthetics were observed, spurned, or merely adjusted to changed circumstances’ (4).

The op.11 piano pieces have a tightly controlled motivic activity full of tonal allusions. The opening phrase of the first, consisting of brief melody, and accompaniment, sets the tone. It is transformed into a concluding phrase for the section while between there
is a three-times repeated two-note phrase (E-G), similarly harmonised with differently
synchronised elements. Perle said: ‘the intervallic cell creates a larger, thematically
significant pattern’ (5), forming melodic and harmonic groupings of minor third and
major seventh. It returns at the end in a manner recalling ternary form. Part of the
accompanying material (major third-minor third-two semitones) is reused as a motif in
the section marked ‘sehr langsam’, underpinned by a three-note figure with tonal
implications, the type of activity identified by Will Ogdon (6). Reinhold Brinkmann
clearly noted tonal centres in this piece (7), but, in complete contrast Allen Forte made
clear that he recognised a different form of activity. Disregarding the strong harmonic
movement of bass-lines and the tonal implications of the melodic material, he refute
any tonal implications in his analysis of the piece in terms of hexachords, pentachords,
tetrachords and their complements, apparently confirming Schoenberg’s break with
tonality (8). Samson’s objections to part of Forte’s method are made clear: ‘the
unpardonable aspect of his theory is its dismissal of those relationships - conscious or
unconscious - which are already audible and significant for the listener’ (9). The second
piece again uses clearly audible motifs that are repeated and transformed. A constant
point of reference is the regularly recurring ostinato which with its rocking triplet minor
third gives a strong tonal feeling. Both Brinkmann (10) and Samson (11) support this
point. It creates a number of static tonal centres, again within a strong ternary structure.
The erratic and volatile third piece is less concerned with audible motifs of the romantic
type and with traditional formal structures. Schoenberg had reduced his reliance on
melodic shapes so much that Perle pointed to the successive thirds as a linking feature
in this piece. Nevertheless, there is a considerable reliance on harmonic movement that
has a strong tonal feel. By removing certain octave displacements in the opening bass-
line, configurations with tonal implications become apparent. The originality of the
piece lies in its formal novelty. There is no difficulty in understanding that the op.11
pieces as a whole are very ambiguous tonally, and can be analysed in these terms, but
they can also be understood in terms of pitch set classifications.

The broad formal shapes still apparent in the op.11 pieces are missing from the brief
and terse Sechs kleine Klavierstücke op.19. Their miniature scale is the most obvious
difference from the earlier pieces. Literal repetition and motivic development are now
much less common or completely absent. The connections with tonality are
considerably more tenuous than those of op.11. Those explored by Kenneth Hicken
(12) are almost totally hidden from our hearing. The first piece uses numerous small
motifs which do not repeat themselves, but interact almost instinctively. Like the third
of the op.11 pieces, repeated thirds are a unifying factor in the second piece of op.19,
but here they are in a microscopic form. Like the ostinatos of op.11 no.2, they do define
a quasi-tonal area but the contrast to other dyads is scarcely of the same order. The
powerful progression that opens op.19 no.3 has a tonal feel, especially in the powerful
bass-line in doubled octaves, even if its dynamics were suppressed by the composer.
There are some near repeats of motifs. In pieces 4 and 5 there are some ambiguous hints
of tonality. The fourth piece is a kaleidoscope of brief ideas that show little connection,
its episodic form typical of the new plans the composer was now using. Some melodic
aspects appear in no.5, but here the rhythmic repetitions of four or five notes and a
number of major and minor thirds towards the end are all that one can seize on. The last
piece is significant for its use of three-note chords, and certainly a piece that is very
difficult to analyse in tonal terms. Forte’s analysis of this piece is convincing because
the chords do not take on a melodic-thematic significance (13).

The same contradictions appear within and between the relatively expansive Fünf
Orchesterstücke op.16 and the terse Three Pieces for Chamber Orchestra of 1910. The
former has formal reprises and repetitions, a strong motivic interest, and hints of
tonality, but the latter have little repetition or development, and no identifiable tonal
sense. The important motifs of op.16 nos. 1 and 5 can be explained in traditional terms
of tonal or quasi-tonal relationships and they certainly sound that way. ‘They have
organic structures, recognisable themes, an overall key-centre and harmonies that do, in
their fashion, direct and punctuate the flow of events’ (14). Forte, on the other hand,
analyses the motifs of the first piece in terms of pitch-class sets (15). In a sense both are correct. In the third piece the static ostinato-like sounds hover around the note C, but the way the chords are used suggest that Schoenberg was now thinking in terms of groupings that are easily identified with Forte’s analysis. Like the piano pieces op.11 the work can be interpreted both ways. The remnants of tonality are present, the motivic structure, except perhaps in nos.3 and 5, is strong and related to tradition. Formal structures can be understood in traditional terms.

To move to the Three Pieces for chamber orchestra of 1910 is an amazing leap, and very similar to the progression from the piano pieces op.11 to those of op.19. In both cases the two pairs of works are separated by only a year. The Three Pieces last less than a minute each (the last is unfinished), with the melodic parts mostly very short and inconsequential. Longer ones are smothered in a web of motifs. A tonal cadence may be construed at the end of the first piece, but it stretches credibility. The third piece, based on a six-note harmonium chord, has a wide range of ostinatos, but there is no formal structural significance as in op.11 no.2. This miniature way of thinking was clearly a new departure for the composer and one that influenced his pupils Berg and Webern.

We can observe the same kind of dichotomy in the works for solo voice and piano or ensemble. The song-cycle of 1908-9 of settings by Stefan George entitled Das Buch der hängenden Gärten was a considerable move forward from the String Quartet No.2. The composer wrote in his essay ‘How One Becomes Lonely’: ‘these songs showed a style quite different from anything I had written before” (16). Later in ‘My Evolution’ he made the conflict clear: ‘this music was distinctly a product of evolution, and no more revolutionary than any other development in the history of music (17). Let us examine these points in greater detail. There is in the cycle a motivic network of the type used in Drei Klavierstücke op.11, with regular appearances of connections to traditional tonality. From the first piano phrases one is aware of an expressive use of the contrast of major and minor thirds. Semitones can clearly be heard as part of this contrast while whole tones and fourths are also encountered. The vocal line is particularly strong in this respect, with implied harmonies at many points in the work. Lessem grouped the motifs very neatly into four cell types (18) and Ehrenforth produced a chart of motifs with some different results (19). They are very short and they only suggest some sort of fleeting tonality, but this is audible. As for the chordal-harmonic implications of the music, H. H. Stückenschmidt observed: ‘the sonic effect of this music is still full of apparently tonal moments, and in some chord-sequences there are still memories of a tonal cadence’ (20). He argues that while no overall tonal centre can be deduced, the constant tonal references have an expressive purpose. Lessem goes further and argues that there is a return to a defined tonal centre in some of the songs (21). In an extensive study of the work, Albrecht Dümling considered the ambiguity involved (22). Taking a completely different approach to the songs’ tonality Allen Forte uses examples of the motifs (23) and of the chords (24) as representatives of pitch-class sets. This clearly is in some form of conflict with Stückenschmidt, Samson and Lessem. To say that both sides are right is to give some idea of the ambiguity of the work; it is certainly not an abrogation of intellectual integrity. Formally we have the same problems. Just as the first two of the piano pieces op.11 have residual features of ternary form many of the fifteen songs of the cycle have a similar connection. Lessem (25) finds these in varying degrees virtually throughout the whole cycle, with ten showing a fairly full or partial ternary structure. Stückenschmidt is more ambiguous: ‘the majority are pure imaginative forms, in which the utterly logical effect and unforced appearance of the structure seem almost a puzzling phenomenon’ (26).

When we turn to the much briefer and angular songs of Pierrot Lunaire, composed in 1912, we are in a different world. Not only is the character of the words different, but the music also operates differently. The move is from Lieder to cabaret. Schoenberg’s use of motifs, tonality and form now has a new perspective. The composer said that the work: ‘gave me a great success by the novelty which it offered in so many respects’ (27). Motivically it is a very concise work. Rosen reports: ‘A short motif of from three
to fourteen notes provides the nucleus not only for the entire melodic development but for almost every note of the accompaniment as well' (28). Yet it is not as simple as that. Lessem identified three groups of three-note cells and assigned plausible meanings to them (29) without any specific tonal connection. He isolated a further new feature in a ‘Pierrot’ motif in numbers 1, 2, 3, 9, 13, 14 and 21 (30), but qualified this by saying: ‘the many variants choose their pitches very freely, the only stable referential factor being overall contour and rhythmic profile’ (31). On tonality, he puts his finger on the difference: ‘tonality, while observable as a subject of ironic reference here and there, is no longer demonstrably operative in the work.’ Any tonal reference comes from outside the work, as happens in Die glückliche Hand, as opposed to the earlier Erwartung which incorporates its tonal quotation fully within its harmonic and motivic structure. The poems of Pierrot Lunaire are uniformly structured with a return to the opening lines which invites some form of recapitulation, but Schoenberg studiously avoided anything like the audibly recognisable reprises found in the piano pieces op.11 and Das Buch der hängenden Gärten. Most of these can only be found by the listener or analyst by very careful study of the contrapuntal textures. He made much less use of chordal harmonies, as shown by Forte’s dearth of chordal examples drawn from this work (32).

The same kinds of conflicts appear in the operas. Although Erwartung is considered one of the pinnacles of atonal opera, it has at its heart the quotation and reuse of the composer’s own song Am Wegrand (‘By the Roadside’) of 1905, in the key of D minor (33). Typically the work is described as unanalysable. I reject MacDonald’s glib dismissal of various approaches to the music (34): they do help us to approach the complexity of Schoenberg’s mental and creative processes and the incredibly complex combination of techniques that he brought to bear on the composition of the work. In a sense the work arises out of a conflict between tonality and atonality and another between long melodic lines and short melodic cells. It is central to the understanding of Schoenberg’s move into new worlds of sound. Formally the work is clear enough, three short scenes and a fourth that divides into a number of parts, but unlike the other works considered in this study, the music is continuous for nearly thirty minutes. Certain sound complexes are characteristic of the work. Rosen (35) abstracted a number of six-note chords, as well as three- and four-note groupings which are performed both melodically and harmonically. This is exactly what Lessem found (36) and what is comprehensively documented by Laborda (37) and Mauser (38). Just like the George songs of 1908-9, the small motif plays a fundamental part in generating the musical material. The conflict in Erwartung is found as mentioned above in the integration of the song Am Wegrand into the texture at the point of climax (39). The significance of this is not difficult to deduce: the roadside is where the woman finds her dead lover and perhaps straining the meaning a little further, the dead man represents tonality as the composer inherited it. The conflict is very clear.

The music of the later opera, Die glückliche Hand, is tense and compressed, using motifs in a way that makes any tonal implications almost totally meaningless. When this work was composed Schoenberg had passed the point of no return. Just as in Pierrot Lunaire, any tonal references came from outside the work. The appearances of the ‘popular’ music in the opening and final scenes are intended to be ironic and simply dramatic in their purpose. The duration is less than twenty minutes and the formal plan is very much more sectional than that of Erwartung.

The conclusion that one can draw from this study is clear. The first of each pair of atonal works moved into the area known as atonality but kept the tonal option open in some way or other. In some cases it was achieved by the formal plans and their implied connections with tonality. Some analysts, however, considered these works to be completely divorced from tonality. As we have seen there is considerable ambiguity in the first group of these works. In the second of the pairs this ambiguity was removed or at the very least reduced to virtual insignificance. It then becomes possible or even necessary to evaluate these works by other criteria. Others can now use this information to make an assessment of the state of the composer’s mind during these critical years.
Notes

2. Ibid., 49-50
4. Schoenberg: op. cit., 86
5. George Perle: *Serial Composition and Atonality* (London: Faber, 2/1968), 20
10. Brinkmann: op. cit., 97
11. Samson: op. cit., 163
15. Forte: op. cit., 11
16. Schoenberg: op. cit., 49
17. Ibid, 86
21. Lessem: op. cit., 57
23. Forte: op. cit., 34 and 63
24. Ibid, 1 and 9
25. Lessem: op. cit., 57
26. Stückenschmidt: op. cit., 46
27. Schoenberg: op. cit., 51
29. Lessem: op. cit., 137-38
30. Ibid, 132
31. Ibid, 132
32. Forte: op. cit., 39
34. MacDonald: op. cit., 76-77
35. Rosen: op. cit., 51
36. Lessem: op. cit., 77-78
38. Siegfried Mauser: *Das expressionistische Musiktheater der Wiener Schule* (Regensburg: Bosse Verlag, 1982)
Enoncé musical et mode(s) de structuration du *pitch space*, ou :
de la relation composition / cognition dans un fragment de l'opus 28, I de Webern

Jean-Louis LELEU (Université de Nice)

Lorsque nous écoutons de la musique, cette écoute n'est pas seulement régie par des lois générales qui tiendraient à la nature même de notre activité cognitive : elle est également dépendante de modèles d’analyse et d’interprétation, et plus profondément de schémas de pensée acquis, éminemment relatifs, dont l'étude relève autant de la sociologie de la culture que de la psychologie. Et cette dépendance ne touche pas moins le musicien averti, le « spécialiste », que le simple mélomane ou le « profane » : le théoricien lui-même, ou en psychologie l'expérimentateur ne sont nullement à l'abri de ses effets. À l'inverse, toute musique qui compte ne manque jamais, dans notre culture du moins, de s'affranchir d'une manière ou d'une autre de tels schémas de pensée, et d'élaborer ses propres modèles, y compris sur le plan cognitif : toute création artistique suppose cette forme d'émancipation.

L'écart par rapport aux schémas préexistants peut s'accomplir au sein d'un idiome musical donné – par exemple le système tonal –, sans que soient remis en cause le mode de structuration de l'espace sonore qui le caractérise, ni, par conséquent, la forme de logique musicale, et le modèle cognitif, qu'implique ce dernier (1). Les choses se consent lorsque l'idiome lui-même se transforme profondément – comme chez les musiciens de l'École de Vienne –, mais également lorsqu'il s'opère un changement de *paradigme* du type de celui que l'on observe à partir des années 1890 chez Debussy notamment, et qu'il illustre de façon comparable ensuite la musique de compositeurs aussi différents que Schönberg, Stravinsky, Bartók, etc. : à savoir la mise en question radicale de cette règle, centrale dans le cas de la tonalité, qui veut que toute musique, dans son organisation propre, renvoie de façon univoque à un mode de structuration précis du *pitch space*, définissant un idiome musical donné. Les manifestations d'une telle mutation, et ses incidences sur le plan de la cognition, sont particulièrement intéressantes à observer dans le cas de la musique dite «atonale», où est exprèsment posé le problème de la structuration d'un espace devenu, selon l'expression de Lerdahl, un espace « plat » (2), l'échelle chromatique sur laquelle il repose étant dépourvue de qualités propres, et n'offrant, de ce fait, aucun moyen de différenciation. Les compositions dodecaphoniques de Webern, tout spécialement, témoignent du souci extrême de parvenir à un contrôle précis et rigoureux de l'ensemble des relations de hauteurs opérées au sein de cet espace, notamment par la mise en place de « dispositifs » destinés à assurer, au moyen de la série, une distribution *symétrique* de toutes les notes autour d'axes ou de *centres* de symétrie bien définis. Mais en même temps, l'énoncé musical, tel qu'il s'articule concrètement – dans sa réalité vivante –, présente dans cette musique nombre de configurations qui sonnent tonallement, c'est à dire font référence, du fait de leur nature même et de leur disposition, à un mode de structuration du *pitch space* basé sur une division *asymétrique* de l'octave, où domine hiérarchiquement la quinte (3). Le propos du présent article est de montrer à partir d'un exemple précis, tiré du 1er mouvement du *Quatuor* opus 28 – à savoir la 4ème des six variations que comporte ce mouvement (mes. 66-79) (4) –, comment est cultivée par le compositeur cette forme d'ambivalence, de *Mehrdeutigkeit*, et les conséquences qu'elle entraîne du point de vue de la cognition.

Il est nécessaire, si l'on veut rendre compte de la logique musicale à laquelle obéit le déroulement de cette variation, de percer à jour le dispositif qui en commande l'agencement : « das gesetzliche Kalkül », selon l'expression de Hölderlin que cite précisément Webern dans l'une de ses dernières lettres à Willi Reich. Ce dispositif est largement déterminé par deux caractéristiques propres à la série utilisée dans le *Quatuor* opus 28. I) Cette série se compose de trois tétracordes de même structure, qui « ordonnent » les notes du *pc set* 4-1 de la nomenclature de Forte – [0, 1, 2, 3] – de façon à produire le motif BACH ; si l'on prend la forme première de la série partant de *sib* (P10) – qui est précisément celle par laquelle commence notre fragment –,
telle que Webern la note dans sa Reihentabelle (5) [ex. 1], le tétracorde a est ce motif BACH lui-même, le tétracorde c sa transposition à la 6te mineure supérieure, tandis que le tétracorde b est à la fois l'inversion et le rétrograde (transposé à la 3ce majeure) de a et de c ; « à la fois », car a, b et c ont cette propriété d'être RI-symmetrical, c'est-à-dire identiques à leur propre inversion rétrograde (transposée), et symétriques autour d'un centre, représenté ici par leurs deux notes extrêmes (les trois centres de symétrie successifs décrivant un cycle de 3ces majeures). 2) La même propriété se transmet, du fait de cette disposition, à la série tout entière : P10 se confond avec R7, et les 12 notes s'y distribuent autour d'un centre de symétrie qui n'est autre que celui du tétracorde b.— Dans le passage qui nous occupe, quatre formes de la série sont enchaînées, de telle manière que le dernier tétracorde de chaque forme est également le premier de la suivante ; on a donc au total la succession suivante : a1 b1 c1=a2 b2 c2=a3 b3 c3=a4 b4 c4 (soit P10 < P6 < P2 < P10). La relation entre les tétracordes ne se modifiant pas (le cycle de 3ces majeures formé par les centres de symétrie de a, de b et de c=a se reproduit indéfiniment), ceux-ci restent d'un bout à l'autre composés des mêmes notes, et l'enchâinement fait en outre revenir à la fin la forme de départ. Mais le plus important est que la succession des 36 notes possède à son tour, dans sa globalité, la propriété d'être RI-symmetrical : elle se confond avec sa propre inversion rétrograde (partant de sol), et s'ordonne tout entière autour d'un centre de symétrie qui est, cette fois, celui du tétracorde central de tout l'agencement (c2=a3). Il y a donc homologie de structure entre les unités des trois niveaux : la microsérie que constitue le tétracorde, la série proprement dite, et la macrosérie qui résulte de l'enchâinement des quatre formes. Celle-ci, de surcroît, se subdivise elle-même en trois unités de tétracordes, dont la première et la dernière coïncident avec le déroulement de P10, tandis que l'unité centrale, qui regroupe b2, c2=a3 et b3, forme ce que Milton Babbitt nomme une « série secondaire » (secondary set) (6), dont les caractéristiques sont identiques à celles de la série principale ; cette série secondaire apparaît également de part et d'autre de l'unité centrale (b1 c1=a2 b2, et b3 c3=a4 b4), le nombre des structures RI-symmetrical présentes dans la succession des 36 notes s'élève finalement à 17 : 9 tétracordes, 4 formes de la série principale, 3 formes de la série secondaire, et l'enchâinement tout entier.

Le « dispositif », cependant, ne se limite pas à ces 36 notes : la variation revêt en effet la forme d'un canon par mouvement semblable, à la 3ce mineure ; à l'enchâinement qui vient d'être décrit s'en superpose donc un second, de même structure, mais dont les tétracordes constituent, étant donné la note de départ, sont différents des premiers (P7 < P3 < P11 < P7). Pourquoi le choix de la 3ce mineure? Cet intervalle, présent au centre de a, de b et de c, est en vérité celui que forment les notes extrêmes du segment chromatique auquel se ramène, en tant que pitch-class set, chaque tétracorde. Deux tétracordes a, b ou c distants d'une 3ce mineure ont ainsi une note en commun, et dans le double tétracorde qui résulte de leur combinaison, cette note commune devient le centre autour duquel s'ordonne symétriquement toute l'unité ; en d'autres termes, chaque double tétracorde forme à son tour une structure RI-symmetrical. Cette symétrie est très clairement réalisée et le moment est ici venu de considérer les hauteurs réelles de l'extrait, par exemple, dans b1 ou dans c1=a2 [ex. 2]. Mais tout ce qui a été dit plus haut au sujet de la 1ère « voix » du canon vaut aussi pour le dispositif global. S'y retrouve, en particulier, la même homologie de structure entre, cette fois, les doubles tétracordes, les couples de formes sérielles (de la série principale et de la série secondaire), et finalement la double macrosérie qui constitue tout l'agencement : les 72 notes de cette macrostructure se distribuent symétriquement autour d'un do#4 « idéal » que représente, en son centre, le do#4 de c2=a3 (7), le même do#4, au centre de b1 et de b4, remplissant la même fonction au sein des unités de 3 doubles tétracordes correspondant à la combinaison de P10 et de P7. Le canon par mouvement semblable (à la 3ce mineure) se double ainsi, secrètement, d'un canon par mouvement rétrograde et contraire (au triton, à la 3ce majeure ou à la 7ème mineure selon l'unité considérée). Les notes centrales des doubles tétracordes successifs décrivent, de leur côté, un cycle continu de 3ces majeures (ou de 6tes mineures) : b4 pour a1, do#4 pour b1, fa4 pour c1=a2, etc. (8) ; celles de b1, b2, b3 et b4 — au centre des couples de formes de la série principale— forment un cycle clos sur lui-même, disposé de façon strictement symétrique autour de do#4 (do#4, fa4, do#4), qui confère une sorte de respiration interne, en l'articulant, à la structure dont les do#4 de b1, c2=a2 et b4 marquent, de manière figée, le pouvoir unificateur.
Tel est donc le « dispositif », dont nous devons nous demander maintenant comment il se concrétise musicalement. Notons, tout d'abord, que son centre, par rapport auquel s'opère toute la distribution symétrique, coïncide de façon précise avec l'unique césure que comporte la variation (cf., mes. 71s., « poco rit » / « tempo », et le contraste dynamique pp / ff ); la correspondance qui s'établit ainsi, toutefois, n'est en rien une équivalence : l'écriture procède, bien plutôt, à une interprétation libre de la structure, en traduisant sous la forme d'une articulation temporelle, marquant le passage d'une première proposition musicale à une autre, ce qui dans le « dispositif » se présente spatialement comme un système concentré de relations internes. — Il est significatif, d'autre part, qu'à cet endroit de la variation les hauteurs réelles — en l'occurrence les notes des tétracordes b de P3 et de P2, et celles de e2—a3—, se distinguent de façon strictement symétrique autour de do#4 : la symétrie marque donc fortement ici le phénomène musical [ex. 3]. Seule fait exception, dans e2—a3, la dyade mis-sib3 (la seule de toute la structure à avoir pour centre do#) ; mais il est aisé de considérer que dans ce moment de tension — où éclate précisément le forte — les deux notes sont justement haussées d'une octave ; le recours à la symétrie comme principe d'organisation, en effet, n'imprime nullement que la distribution symétrique doive s'effectuer de manière toujours littérale, et pour ainsi dire mécanique : Webern a maintes fois, au sein de figures musicales régies par ce type de structure, fait des déplacements d'octave, en tant qu'écarts par rapport à la disposition stricte, un agent spécifique de l'expressivité (9). Plus problématique, sur le plan de la cognition, est en revanche le statut des figures symétriques elles-mêmes : si auditivement la relation s'établit sans difficulté, de part et d'autre de la césure centrale, entre le tétracorde b de P3 (2nd violon, mes. 71s.) et celui de P2 (alto, mes. 72s.), ce dernier a toutes les chances d'être entendu comme la reprise textuelle, un demi-ton plus bas, du précédent, et non comme son inversion rétrograde ; et cela vaut déjà, au sein même de b1 et de b2, pour les figures du violoncelle et du 2nd violon (mes. 67-69 et 70-72) : l'imitation par mouvement direct et semblable est à l'évidence plus prégnante ici que l'imitation par mouvement rétrograde et contraire. Quant aux relations qui s'établissent entre des tétracordes éloignés — par exemple entre a de P6 (1er violon, mes. 68-70) et e de P3 (violoncelle, mes. 72s.) — dont toutes les notes s'inversent autour de mis5, il est inutile de souligner qu'elles restent des plus secrètes. Pourtant, la structuration de l'espace sonore s'effectue bien, à tous les niveaux du texte musical, selon le principe décrit plus haut ; grâce au dispositif mis en place, un ordre régne, même si l'oreille ne peut en saisir le ressort intime. Tout repose ici sur un calcul et un pari d'ordre à la fois compositionnel et cognitif, qui est de considérer que la symétrie ne manquera pas d'agir sur la perception de façon subliminale, et donc de s'imprimer dans l'image auditive — sans qu'il soit besoin de l'appréhender comme telle. En assignant au do#, dans tout le passage, une hauteur fixe (alors que toutes les autres notes sont mobiles), et en faisant du do#4 le centre de symétrie effectif de b1 et de b4, Webern signale du moins le statut parcellier de cette note : un statut équivalent à celui d'une tonique (10).
conservée dans \( b_4 \);
ces pizz.
se combinent alors avec ceux de \( a=c \) au sein de figures complexes où se détachent avec un relief nouveau les dyades de \( a=c \) jouées arco [ex. 5].

Cet examen des figures rythmiques permet de mieux saisir comment se forment au sein de l'énoncé musical les configurations de hauteurs réelles. Les éléments du dispositif qui contrôle toute la variation sont en effet réinterprétés par l'écriture de manière à produire des configurations de type tonal, c'est-à-dire faisant référence à la gamme majeure/mineure et à sa division asymétrique de l'octave. Dès le début du fragment, le \( sol \) joué par l'alto sur le 1er temps de la mes. 67, après l'anacrouse que forment les deux notes du 1er violon, tend à prendre la valeur d'une tonique, et l'intervalle \( do-fa# \) qui suit immédiatement suggère la 7ème de dominante du ton de \( sol \). Mais surtout : le déroulement de la variation tout entière est rythmé par le retour de plusieurs gestes cadentiels, selon une périodicité due au cycle de 3ces majeures qu'engendre la succession des formes sérielles. Le premier de ces gestes cadentiels implique les deux dernières dyades de \( a=c \) et la 1ère dyade de \( b \) : ainsi, mes. 67, les notes \( do-fa#-si-re# \) composent une harmonie de 9ème mineure, qui se résout sur le 1er temps de la mes. 68 (avec \( mi \) comme fondamentale de l'accord de tonique, et \( la \) comme appoggiature de la tierce) ; le même enchaînement se réitère aux mes. 70 (ton de \( do \), 72s. (ton de \( lab \)) et 74s. (de nouveau ton de \( mi \)). Le deuxième implique la dernière dyade de \( b \) et les trois 1ères notes de \( a=c \) jouées pizzicato : en particulier, le mouvement mélodique \( fa#-sol \) du 2nd violon, à la fin du 1er volet, s'entend comme celui qui porte de la sensible à la tonique (à cette dernière s'adjoing l'au temps fort la tierce \( si \) ; l'effet de résolution qui en découle donne ici tout son sens à la césure. Le même enchaînement, pris cette fois dans le flux du discours musical, se retrouve aux mes. 74 et 76s. (avec, respectivement, \( mib \) et \( si \) comme toniques) ; seule fait exception -à la fois sans doute pour éviter un excès de régularité et pour ménager l'effet de la césure centrale- la figure de la mes. 69, où le \( mib \) de l'alto, joué arco et tenu, sonne comme la 5te de l'accord parfait mineur de \( lab \). Entre ces deux mouvements cadentiels se forme à quatre reprises une harmonie de tons entiers qui, dans le contexte, tend à sonner comme un accord de 6te augmentée -par ex. mes. 68 \( re-sol#-do=si#-fa#- \), tandis que retentit en son milieu la tonique correspondante, qui n'est autre que la note faisant office de centre de symétrie dans chacun des couples de formes sérielles (\( do#4, fa, do#4 \)).

Ces configurations tonales se font et se défont, dans le cours de la variation, selon une logique précise, mais qui ne doit rien aux fonctions tonales classiques : la conduite du discours est ici entièrement déterminée par un dispositif basé, nous l'avons vu, sur l'idée d'un espace sonore où règne la symétrie, et qui requiert cette «perception absolue et unitaire» dont Schoenberg a souligné l'importance (11). Une tension se crée ainsi à l'intérieur du texte musical entre deux modes de structuration du pitch space antithétiques, mais également entre deux formes de temporalité. Car une certaine progression a bien lieu au sein de la variation, qui prend appui sur le rythme même des transpositions successives des couples de formes sérielles de 3ce majeure en 3ce majeure. Or, cette progression, qui peut sembler ouverte, s'effectue dans un laps de temps clos sur lui-même, dont la fin coïncide avec le commencement et revient à lui. De là le caractère évanescent des dernières mesures, où s'interrompt la référence à la tonalité et le dynamisme interne qu'elle impliquait. Etre sensible à cette ambivalence, c'est admettre la possibilité d'un modèle cognitif dans lequel ne soit pas posé le principe d'une structuration homogène de l'espace sonore. Le type de dispositif auquel Webern recourt dans le Quatuor opus 28 serait, en lui-même, de nature à fonder une «hiérarchietonale» conforme aux trois conditions posées par Lerdahl et Jackendoff : on dispose en effet grâce à lui d'une «collection de hauteurs» (les 12 notes de la gamme chromatique), d'une «tonique» en tant que «major point of stability» (ici le centre de symétrie de toute la structure), et de la possibilité de mesurer la «stabilité relative» des autres éléments (= les centres secondaires) en fonction de la distance qui les sépare de cette «tonique» (12). Mais le compositeur n'exploite pas dans cet esprit le principe d'organisation dont il se dote : en misant sur les tensions qui naissent de la surimpression, et donc du conflit de deux idiomes —celui qu'il se forge, et celui qu'il reçoit de la tradition—, il dénonce d'embellée, et rend caduque, toute tentative de plier le message musical aux exigences d'un modèle strictement linguistique.
Notes

(7) Sur les 36 dyades qui se forment à l'intérieur de la structure globale, 23 –comme le montre l'ex. 3, où le 2ème enchaînement est présenté sous forme rétrograde– ont pour centre de symétrie effectif do#: dans les autres cas, l'inversion se fait en raison de déplacements d'octave autour de sol, sol# ou do. 
(8) Pour c3=as1, le centre de symétrie autour duquel s'agence l'unité est, du fait du dédoublement du la, mib. 
(9) Cf. notamment les Lieder op. 23 et op. 25 (par ex., dans l'op. 23 n° 1, mes. 11, «das an die Tzen rührt»).
(10) La fonction de tels centres de symétrie (que l'on trouve surtout dans les opus 28 à 30) est analogue à celle des axes de symétrie (liés à la combinaison de formes P/I) autour desquels s'ordonnent également chez Webern maintes sections, voire des mouvements entiers (cf. notamment les opus 21, 22 et 27). Si la présence des seconds a été fréquemment commentée par les musicologues, celle des premiers est semble-t-il restée jusqu'à ce jour inaperçue. 
Nature, Ethnoaesthetics—Turangalîla: Messiaen’s analogously implementing exotic sound material in his symphony

Alexandra Hettergott
Institute of Musicology, University of Vienna
Universitätsstrasse 7, A-1010 Vienna, Austria

Introduction

The paper on the one hand treats the Turangalîla-symphony as a whole, regarding basic idea and conception, on the other hand analyzes the aspects of influence and structure: Messiaen’s being inspired by the gamelan’s exotic sound colour and rhythm as well as by the manifold melodic patterns of bird song. While the background to history of origin and idea of Turangalîla refers to Messiaen’s universal conception of nature, both compositional figures and orchestra colouring can be related to their prototypical model: like the gamelan as regards struck metal instruments, rhythmical organization and similarity of the resulting overall sound impression, or like the bird songs in their melodic patterns, in a stylized manner assigned to solo piano. Thus, first it is shown in which way sound and rhythmical character of Messiaen’s gamelan are related to the original Javanese and Balinese gamelan (IIIRD), and secondly how and to what extent the compositional figures do reflect the melodic patterns of the three birds they are based on (VIIITH movement).

Together with the “song of love and death,” Harawi (1945), for voice and piano, and the Cinq Rechants (1949) for twelve unaccompanied mixed voices, the symphony named “Turangalîla” (1946-8), actually a Sanskrit word, whose meanings in Messiaen’s interpretation include all at once “chant d’amour, hymne à la joie, temps, mouvement, rythme, vie et mort,” is the middle part of a trilogy on Tristan and Yseult. The symphony itself processes gamelan percussion and Hindu rhythmic formulae, as well as the stylized and idealized songs of nightingale, blackbird, and garden warbler. Here too, as in other compositions of his, Messiaen takes use of his rich repertory of ideas and compositional techniques, combines single originally distant elements with the total effect being pure non-conform Messiaen. And here, too, his transcoding of nature and ethno sound and aesthetic expression pays homage to the grandeur of nature as a revelation of God’s majesty, an intention which too, in effect, however his being completely devoted to a Roman-Catholic faith, bestows a certain pantheistic aspect on his oeuvres. Also the gamelan, originally used in a ritual context, as for religious dance (as the Pendet example examined below), does fit in this context. The quasi-gamelan used here does actually form a small orchestra within the large orchestra, in addition to piano calling for celesta, vibraphone, glockenspiel and other metal percussion, which together with the ondes martenot (an early monophonic electronical keyboard instrument, constructed in 1928 by E.Martenot) gives rise to a Balinese orchestra colouring. The symphony’s rhythmic diversification in “rhythmic characters” and “non-retrogradable rhythms” intendedly finds its correspondence in the original gamelan instruments which can be divided into four main groups: some playing the basic melody (saron, a metallophone provided with a box resonator), others decorating or contrasting the melody (bonang, a set of horizontally suspended metal gongs; and gender, a metallophone with bars resting on ropes each provided with a hanging bamboo resonator), punctuating melodic phrases (gong), signaling rhythm, style, and tempo of the music (kendang, a two-headed cylindrical drum). Typical of the IIIRD of the ten movements, Turangalîla 1, is its being structured in several compositional layers: while in one layer being allocated to three different rhythmic characters—“mineral timbre” (maracas), “vegetable, wooden timbre” (wood block), and “animal timbre” (bass drum)—in another one amongst others it does combine ondes martenot (showing a metallized echo timbre), bell, vibraphone, clarinet, and double bass in a first (cf. Fig. 1), and the gamelan of celesta, glockenspiel, vibraphone, piano in a second theme (cf. Fig. 5).
Fig. 1. Successively introducing several (metall) instruments at the beginning of the IIIrd movement, Turangalîla I, in spatial distribution (cl = clarinet, OM = ondes martenot, bl = bell, vb = vibraphone).

A – Metall gamelan sound

Example 1: The first theme of the IIIrd movement (Fig. 1) does begin by successively introducing several intercoursing (OM, cl) and punctuating (bl, vb, db) instruments. Provided with a resonant gong placed in the diffuser the ondes martenot will later enrich the artificial gamelan with a special metallic sound, getting the closer adapted to the metallic shimmering and slightly beating sound of the Balinese gamelan resulting of several of its instruments, as the gender, playing in pairs with the “female” one being slightly lower tuned than the upper “male” one.
Example 2: The introducing section is followed by a very dense one corresponding to the gamelan in both rhythmic and metallic character (mm. 24-41, cf. Fig.5), a subsequent transitional passage where in addition to the ondes martenot the small flute makes itself heard with a descending tune (reminding on a white-tailed eagle; mm. 42-54), and again another one in which finally the different timbres assigned to the interacting "rhythmic characters" become dominant (mm. 55-66, cf. Fig.6).

Messiaen, Turangalïla: mm.24-66 (nos.[2]-[7])

Example 3: The gamelan and metall character is yielded by providing keyed percussion instruments similar in timbre, comprising glockenspiel, celesta, and vibraphone, the latter in correspondence to the original gender metallophone (Fig. 3), however the gender's bars not being tuned harmonically (ie., no 2:1 relations).

---

Fig.2. Messiaen, Turangalïla I, mm.24–66 (nos.[2]–[7]).

---

Fig.3. Gender–C♯5 vs. vibraphone–C♯5; note that the interval (Δf) between the gender's first two partials is about VIII plus IV.
The residual pitch perceived from the gamelan’s melodic patterns—the example shows an excerpt from the religious Balinese Pendet dance, as grossly given in the transcriptions below—has its striking correspondence in the analyzed IIIrd movement: while in mm.24ff. (no. [2]; cf. Fig. 5) the artificial gamelan is structured in six chords per measure, but highly corresponding in pitch (eg., in m. 26: C – A♯ – A – G♯ – G – E), tempo (MM. eighth = 100), and meter (3/8), with meter 2/4, same tempo, and set with eight/measure the correspondence becomes even closer in mm.73ff. (no. [9]; eg. in m. 75: C – A♯ – A – G♯ – G – E – G♯ – A), with the originally non-distinctive pitch height being reflected by the preceding grace notes. In the example below the basic line is given by the sarons, the very first tone, too, by a gong, whereas the upper paraphrasing line is played by pairs of genders and bonangs. In these three segments also a certain figure-background effect does become apparent: while in segment 1 the upper melodic line is prevailing (+B), in segment 3 it is the lower one which becomes the more dominant (+A).

Gamelan (Bali): Pendet (segments 1-3)

Fig. 4. Segments from the Balinese Pendet dance; note that the percussive patterns and the figure-background effect can also be seen in the RMS-track given above the spectrogram.

Segment 1 (J = 120)

Segment 2 (J = 112)

Segment 3 (J = 100)
Messiaen, Turangalîla 1: mm.24–41 (nos.[2]/[3])

Fig.5. Messiaen, Turangalîla 1, mm.24–41. The spectrogram shows, too, a structural intertwining pattern resulting of the ondes martenot's descending, and the violoncelli's ascending glissando. (It is furthermore focused by using the Bark scale stretching the lower frequency range in correspondence to its being better resolved by human audition.)

B – “Rhythmic characters” and “non-retrogradable rhythms”

Beginning in about the middle of the movement (mm.55ff., no.[6]) a further rhythmic theme becomes apparent, three rhythmic characters—increasing, decreasing, staying—assigned to “mineral” (maracas), “animal” (bass drum), “vegetable” timbres (wood block), respectively. That is, the bass drum as the “active,” augmenting, the maracas as the “passive,” diminishing, and the woodblock as the “indifferent,” non-changing character all at the same time provided with non-retrogradable (= mirrored, as a cipher for neutralized time) rhythms, as in mm.: bass drum maracas woodblock

<table>
<thead>
<tr>
<th></th>
<th>bass drum</th>
<th>maracas</th>
<th>woodblock</th>
</tr>
</thead>
<tbody>
<tr>
<td>61</td>
<td>± 66</td>
<td>61 ± 66</td>
<td></td>
</tr>
<tr>
<td>66/65</td>
<td>± 61/62</td>
<td>62 ± 65</td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>± 64</td>
<td>63 ± 64</td>
<td></td>
</tr>
</tbody>
</table>

Messiaen, Turangalîla 1: mm.55–66 (nos.[6]/[7])

Fig.6. Messiaen, Turangalîla 1, mm.55–66 (nos.[6]/[7]).

The theme, originally being solely rhythmic, therefore makes itself furthermore noticeable by means of different timbres provided with a certain “lively” meaning.
The VIth movement, *Jardin du sommeil d'amour*, depicts the landscape of the two lovers' garden with them transfiguredly asleep outside time, flooded with light and flowers—culminatingly expressed by the songs of nightingale (*Luscinia megarhynchos*), blackbird (*Turdus merula*), and garden warbler (*Sylvia borin*) introduced by piano. In contrast to several of his later oeuvres, as *Oiseaux exotiques* for small orchestra (1956), in this movement the birds' songs are less originally depicted but a means to yield a certain style of expression. Consequently, the melodic patterns are highly stretched in tempo as compared to the original ones, for the music depicts a transfigured and quite non-earthly scene, here and now free from being constricted in both time and space. However, Messiaen carefully translates the particular bird's singing style: the blackbird's de- and ascending, in- and decreasing melodious verses in a lower range and slower manner; the nightingale's rich repertory of sound patterns, repeating one-tone ones as well as over a wide tonal and temporal range varying ascending/descending, accelerando and ritardando tone series, often in imitation of other birds' sound elements; the garden warbler's very fast and irregularly melismatic tunes. Accordingly, the piano adorns the movement's orchestration with diverse figures demanding quite an extreme technique, in simultaneous arpeggios, chord cascades, and combinations of extremely high and low registers in both hands. Besides the piano part, here too the ondes martenot play an important role using the "palm" producing sympathetic vibrations as a special diffuser, conveying the leitmotif-like "love theme."

Set with six for strings and ondes martenot, and with no accidentals for wind instruments (flute, clarinet), percussion and keyed instruments, the score furthermore gives rise to the light-and-shadow setting of the VIth movement—a phantastic, even surrealistic scenery, like in paintings by Max Ernst (the birds!) or Marc Chagall, with the two lovers going beyond themselves, being in complete (and colorful) mysterious transfiguration expressed in music.

References


Recordings

THE RISE AND FALL OF THE EXPERIMENTAL STYLE OF THE BEATLES
The Life Span of Stylistic Periods in Music

Tuomas Eerola, Graduate student,
University of Jyväskylä, Finland

INTRODUCTION

The knowledge of the musical styles is essential for the analysis and appreciation of the musical works. This knowledge is "encoded" into the different stylistic features which the listener inherently knows if the listener is familiar with the style. Musical styles are always evolving and known to have variable life spans which are often implicitly known to have an organic development, a "rise-and-fall" pattern. Though reasoning such as this is often based on the nineteenth century ideology, it might plausibly outline historical processes at a certain level, which might be due to the way we categorize data about the historical change. Robert Gjerdingen (1988) demonstrated this by an extensive research on a mid-level musical schema and its chronological distribution but the results in this area need to be replicated using several stylistic features. Therefore, the goal of the present study was to apply the model of the distribution of the features to a different type of music from a different period and even to a different time range.

The Beatles is a fine example of a band that had achieved a major stylistic development in their music. Literature that deals with the Beatles often summarizes their career as a rise and fall. Many novel ideas were realized by the Beatles in 1965-68, which has been named their experimental period. The early period, 1962-65, is the foundation against which the new ideas are evident and the late period was between 1968-70, analyzing them both enables to show how consecutive style periods work. Their stylistic change is apparent and well-known and has also been interest of musicological writing. Also, a question only briefly discussed in this study will be why the novel ideas were invented in the first place, where they came from and what influenced their use. The aim in this work, however, is mainly to verify the shape, e. g., the statistical distribution of stylistic features, of one style period.

THEORETICAL FOUNDATION

There are plenty of examples how consecutive styles have lived and died to act as general basis for historical processes in Western tradition. The following terms clarify the theoretical aspects used here:

Style (periods) are arranged by the uniform traits, which either alone or combined in variable ways form stylistic features. These are perceived as meaningful units (schemata in cognitive psychology), taking into account the experience of the listener. Stylistic features can consist of multiple parameters which create together meaningful units. Here
these are based on the musicological literature about the Beatles and on the comments the songwriter themselves have said.

Theories of style change point out the external constraints, first of all caused by ideology and thus affecting the parameters of action (politics, economics, social arrangements), and internal constraints, as a shift in the styles of thinking. Consequently, expectations and breaking them has been explained as the reason of change, both by ethnomusicologists and cognitive musicologists. At a larger level the change is seen as an alternation between a period of stability and a period of revolution - convention and invention, reminding of a Kuhnian process of scientific progress. Although the process is a continuous one, historians attempt to divide history into distinct periods and the results are the style periods we know, ranging from one composer’s singular style period to epochs that can last for centuries. However, there is a discrepancy between the way periods are distinguished as solid blocks and the way organic development is used to describe the gradual change. Gjerdingen's model (1988), which depicts concretely the abstract life span of one musical period, demonstrates more satisfactorily the nature of change at a historical level.

Gjerdingen's hypothesis was that the population of a stylistic feature across time approximates a normal, bell-curved statistical distribution. However, the distribution curve is asymmetrically distributed, which is as he claims, because there is a factor modifying it - the memory. Another hypothesis of his was that a schema will exhibit a curve of typicality (the peak is where a prototypical occurrence is found) similar to its population curve. Gjerdingen presents the normal and the modified normal distribution in a simplified way as seen in Figure 1.

The curve in Figure 1 represents both the population and the typicality of stylistic feature. Previous or older, established schemata inhibit the recognition of new ones, until subsequent realization of the new and re-evaluation of earlier examples, with the effect of a sudden increase in the perceived population. After the peak a similar process affects the descending curve to the opposite direction. His results, termed here as the normative life span of style, are here regarded as a general model.
METHODOLOGY AND MATERIALS

The aim was to study the life span of the experimental period of the Beatles and compare the results with the normative life span. I approached this task in stages:

(1) I defined the musical periods of the Beatles and separated the music of each period into significant stylistic features, relying on the literature on both instances. My aim is not to have a complete definition of the Beatles' music but a sufficient one, covering different kind and level of characteristic features. Therefore most of the basic elements of analysis are included in my material, yet some are missing because of their inherent complexity. The early period was studied by following the next (7) typical features: Cover songs, certain ornament, basic line-up, three-part singing, harmonica, woo and yeah - screams, romantic lyrics, and the experimental style period by (11): Changing meter, flattened VII chord (bVII), tone repetition, descending bassline, static harmony, political lyrics, classical instruments, Indian instruments, effects, nostalgic lyrics and psychedelic lyrics. For the purposes of this study, the late style period consisted of the attributes common to the previous style periods.

(2) I registered every occurrence of the features across the Beatles' whole recording career and compared the resulting statistical distribution to the normative life span. This was done for each feature individually. Selected features of the experimental period were studied in detail to see how the experimental period began (nostalgic lyrics/classical music instruments and psychedelic lyrics/corresponding instrumentation). Nostalgic lyrics were the first major innovation apparent in Beatles' music, - also clearly indicated by themselves - and were adorned with an innovation that is well-known - classical music instruments. Then all the features of each period were displayed in whole.

(3) Then I attempted to relate my results back to the concrete level by using the concept of prototype. I discovered those songs that fulfilled as many features as possible belonging to each period, to learn what would a "typical" song of the period be like.


Statistical analysis

I adapted Gjerdingen's model to fulfill my needs by reconstructing his results in the form better suited to the comparison of the materials. Also, it was essential to be able to divide and arrange the Beatles' recording career in a coherent way to make valid comparisons. I used seven-months periods as the interval, and treated my samples as a relative amount (%/album) and always used the first recording date of the song when assessing their place in the chronology.

Quantifying the data occurred simply by dealing with the features only in the sense if they fulfilled the definitions or not, without ranking or rating them any further, that is, without considering the length or the style of playing. One song can represent several features which may even represent different style periods. This method is equivalent of the one used in Gjerdingen's study.
RESULTS

Displaying the average distribution of all the eleven features of the experimental style of the Beatles and the normative life span of style summarizes the main results (Figure 1).

The curves are surprisingly similar in shape, which was substantiated by the resulting correlation values. Reliability was tested by using split-half method, which also yielded highly significant correlations. The effect of memory was tested by comparing the curves with the normal distribution. None of the curves of the individual features were skewed on the opposite direction.

The early period features were also compared normative life span of style. Most of the correlation values were highly negative. These results were not contradictory to the results obtained previously. Rather, the population curves of the early period features were comparable in their shape but were in a different phase than the population curves of the experimental period and the model: The early period features were already at their peak in 1962, because they were learned earlier, during the formative years (1957-62) when the Beatles imitated the music of their American idols. Therefore the different periods could be seen to overlap each other and the chronological presentation of the features could tell in its own objective way what the periods the Beatles' three periods were like and when they were. The alternation between convention and invention was also apparent. The results were also very compatible with the comments the Beatles have given themselves.

The prototypical songs for the experimental period were (in order) Strawberry Fields Forever, A Day In The Life, Within You Without You, I Am The Walrus and Penny Lane, portraying the period remarkably well (recorded also within the population peak of that period), supported by the common knowledge of those facts. In a way, a person who
has some stylistic knowledge about the Beatles, has probably abstracted a prototype that would include most of the features listed here. Therefore a perfect example, such as *Strawberry Fields Forever*, might be very easily remembered (priming effect). Other periods displayed generally similar results supporting the hypothesis. Nevertheless, there are also other songs that are generally regarded as very typical for the Beatles that failed to show up in the results, but as this study focuses on the periods, some information must be omitted in order to do abstractions such as this.

**SUMMARY**

The rise and fall of the Beatles' experimental style was found to be more than a metaphor: The stylistic features of the Beatles' experimental style and their statistical distribution exhibited curves highly similar to the normative life span. The high degree of the relationship between the model and the material in this study came initially as a surprise. Nonetheless, the results were consistent in subsequent examinations: The results were at first replicated by studying the nostalgic lyrics and classical instrumentation. The initial connection between these features was also considered but they were mainly inspected to describe the way the process of change started. The distribution of the early period features was almost the opposite of the model and the three stylistic periods of the Beatles and the organic nature of style change and the overlap of the periods were also demonstrated. The prototypical songs obtained were appropriate to the common knowledge and the literature about the Beatles and thus reinforced the connection between the typicality and population, the validity of the chosen stylistic features and ultimately the concept of the life span.

However, based on one band and mainly its one stylistic period, it is unwise to generalize the results freely to different musical styles.

Although many variables, such as artistic events, political events or composers' lives also affect the history of a style, the very act of categorizing seems to force upon the data an anticipated shape because we all have minds that work in a similar way. This anticipated shape might be commonly known as a rise and fall, the prototypes directing the evaluation of the periods. The effect of memory should not be understood in general terms (implicit, long-term memory), rather as composer's intention to use new ideas (innovative strategy). Although the methods used in this study do not tell us why the changes happen or where they originate from, it could be a way of illustrating the change, supporting other methods.
The analysis of Roman Catholic Church Music to show the spiritual and socialization potentials of the song.
Elizabeth L. Theobald (research student, King Alfred's College, Winchester)

My research deals with the use of music in Roman Catholic liturgies in England and Wales since the Second Vatican Council (1962-65). Whilst researching this area it has become evident that current methods of evaluating music favour particular styles and forms whilst ignoring the various functions which music fulfils within a liturgy. To explore this fully the investigation of music had to include the social and spiritual settings and contexts as well as investigating any positive and negative impacts which liturgical music has upon the congregation. The impact of the music included attempts, by members of various congregations, at assessing why the music has the impact it did.

Although my research does touch upon the findings of Leonard B. Meyer (1967) his research did not mention the phenomenon of the performing audience; a term used by Bernard Huijbers (1980) to describe the post Vatican II congregation, which is both listener to and performer of the music used in liturgies. Occasionally this performing audience also moves into free improvisation thus encompassing all three aspects of musical communication: composer, performer and listener. The performing audience also requires the investigation of the impact of music not only upon the individual, which Leonard Meyer and Susanne Langer explored, but upon the community as a whole.

When investigating the social construction of knowledge, Peter Berger and Thomas Luckmann (1967) defined the process of socialization as being: the process whereby the world is perceived as a meaningful and social reality. They outlined a difference in the process between that of society and that of an individual. In a society, or a community, the sub-processes of socialization (externalization, objectification and internalization) occur simultaneously. In an individual, by virtue of the chronology of skill acquisition, there is a division of these sub-processes into primary and secondary socialization. In primary socialization the individual uses externalization and objectification. There is identification by the individual with the perceptual pattern of the socializing influences. The whole process is emotionally charged and begins before the acquisition of language, logic and analytical skills. The secondary socialization of the individual is reliant upon linguistic, logic and analytical skills: the atmosphere is less emotionally charged and the individual is a partner in the socialization into an institutional sub-world. Secondary socialization works with objectification and internalization.

Both primary and secondary socialization processes are life-long processes in which the socializing influences change with time. Any perceptual pattern acquired in primary socialization, or learnt in secondary socialization, is subject to maintenance and change mechanisms. A perceptual pattern is only valid whilst it provides meaning for experience. Any perceptual pattern which provides meaning is maintained. A perceptual pattern which fails to provide meaning for an experience is either changed so that the experience is now meaningful or the experience is rejected. These
mechanisms of maintenance and change also exist in the socialization processes of a society or community.

Allied to the primary and secondary socialization processes are the types of spirituality expounded by Gerard W. Hughes (1985) and based on the research into religion of von Hügel. Gerard Hughes suggests three elements of spirituality: institutional, critical and mystical. The institutional element accepts without question what is proposed by the surrounding community: most prevalent is the need to belong within the community. The critical element questions, and may reject, everything and anything. The mystical element is where the emphasis is away from the incoming information and towards the expression of knowledge found within (either that of the individual or the community). Gerard Hughes emphasises that each of these elements co-exists but the balance between them within each individual and community changes with time.

In this way one could say that the institutional element of spirituality utilizes the same sub-processes of socialization as primary socialization. The critical element of spirituality incorporates the sub-processes of secondary socialization with a greater degree of evaluation being evident than was suggested in the secondary socialization process. Finally the mystical element communicates and expresses, one might say objectifies, the subjective internal reality: the aim is not to impose this perceptual pattern upon others but to clarify it for oneself.

With the advent of a concern for pastoral liturgical music the spiritual balance of the music is going to be a large factor in the effectiveness of the music used. On the largest scale of considering music the institutional element is more concerned with the known music, styles and placement within the liturgy. The impact is perceived in terms of tradition, nostalgia and the function of uniting the community rather than any challenge within the words or the music. The critical element is more concerned with the analysis of meaning within the languages being utilised, both word and music, and because of this the detail and structure are important. The mystical element is where the function of the music is to express and explore; the linguistic and logic elements are secondary to the auditory space provided.

These elements are noted by Meyer when he investigates the linguistic perception and pattern recognition when someone listens to music. He proposes that music can be heard to such an extent that it ceases to yield meaning of itself. To this he gives two exceptions:
1) ritual music where rite, words and meaning are fully interlinked within a culture which places importance and meaning within the repetition itself
2) Music which has a high degree of random elements which, he insists, contains much information but communicates little.

Thus that music which yields meaning of and in itself impacts upon the critical element of spirituality. The music with a high degree of random element may also be that music which is simple enough to encourage free improvisation from the community itself.

383
Philip Tagg (1994) describes two culturally perceived areas of the soundscape of music: anaphones and 'figure/ground' which he translates as being 'melody/accompaniment'. Anaphones are defined as 'another way of sounding something that already exists as perceptible sound, movement or touch.' The sonic anaphones are thus musical imagery which is portrayed according to the technology of the time and perceived through a cultural filter. The figure/ground concept is a continuum from polyrhythm to heterophony via a duality of melody and accompaniment in which the accompaniment is perceived as subordinate.

Most of the music written after the Second Vatican Council and used in English and Welsh liturgies has one of the following three forms: strophic, verse and chorus and call and response. Each of these forms exhibits a high degree of what Meyer termed 'redundancy', that is internal repetition, even if only by virtue of having several verses each time it is used. Other forms have proved harder to teach to congregations.

When considering liturgical music in English and Welsh Roman Catholic communities the cultural value placed upon repetition prior to the Second Vatican council is still prevalent in many areas. Over 60% of the practising Catholics of the time went to masses which were entirely said. This is no longer true. My research shows that over 70% of practising Catholics now go to a mass where music is used. Although most masses now do have music 'traditional' hymns are important to those who regard music as being a means of participation and a symbol of unity. This is the institutional element. The critical element is all the linguistic attributes which develop: phrase structure and harmony. Music which impacts upon this element of spirituality uses small scale anaphones in word painting and imitation between the melody and accompaniment. The music which impacts upon the mystical element of spirituality falls into two types:

1) Which Meyer describes as having reached the point where the music has no more information and where the listener daydreams through it and is shown in the comment: 'I found it boring at first but then found that I could pray through it.' This is where the mystical element is exploring.

and

2) Which Meyer termed as having a high degree of random elements and information but a low degree of communication which is where the mystical element is expressing itself in free improvisation. This is most clearly seen in what is termed as 'singing in tongues', popular among some charismatic movements within the Roman Catholic church.

In the mystical element anaphones are large scale and structural whilst the rhythm and melody tend towards the motivic. The melody/accompaniment distinctions are extreme, either very distinct where the melody is of prime importance (for example in plainchant) or intermixed where each line is its own melody (as in Taizé chants and 'singing in tongues').
<table>
<thead>
<tr>
<th>Structure</th>
<th>Institutional</th>
<th>Critical</th>
<th>Mystical</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Known pattern, within already existing musical knowledge (Meyer)</td>
<td>Interrelationships and developmental irregularities which memory smoothes out (Meyer)</td>
<td>Simple or predictable.</td>
</tr>
<tr>
<td>Rhythm</td>
<td>Known pattern.</td>
<td>Developing gradually (Meyer)</td>
<td>Motivic.</td>
</tr>
<tr>
<td>Musical drive</td>
<td>Known pattern</td>
<td>Harmonically driven to the cadence points - language of culture (Meyer)</td>
<td>Melody and rhythm</td>
</tr>
<tr>
<td>Imagery (anaphones -Tagg)</td>
<td>External to the music - symbolism in usage (Meyer)</td>
<td>Small scale: sonic, kinetic or tactile imagery</td>
<td>Structural sonic, kinetic or tactile</td>
</tr>
<tr>
<td>Phrasing</td>
<td>Known pattern</td>
<td>Developmental and longer (Meyer)</td>
<td>Motivic</td>
</tr>
<tr>
<td>Texture</td>
<td>Known pattern</td>
<td>Fully realised discouraging free improvisation by the musically untrained</td>
<td>Thin - inviting improvisation even by the musically untrained</td>
</tr>
</tbody>
</table>

Perceptual accounts of the structure of electroacoustic sounds tend to assume that, firstly, timbre is the issue with which electroacoustic musicians grapple, and secondly, an understanding of timbre derived from instrumental acoustics and psychoacoustics provides a starting point for building a psychological understanding of how we listen to electroacoustic sounds. Notions of timbral hierarchy (Lerdahl, 1987) or timbre space (Wessel, 1979) whilst attractive in the way in which they suggest a continuity between structure in tonal music and structure in ‘timbral’ music, ignore many of the more interesting structural possibilities within electroacoustic music. Some of these possibilities exist in the way in which elements in an electroacoustic piece relate one-to-another, i.e. intrinsically, some in the way in which these elements relate to events outside the piece, i.e. extrinsically. The kinds of relationships a ‘timbral’ approach fails to capture are those which are constrained by and generated from our familiarity with the origins of particular sounds, and moreover, the lawful nature of the relationship between acoustic events (sounds) and their mechanical causes (origins). Such views of sound and its direct connections with causation are commonplace in the writings of some composers (Smalley, 1986; 1992; Wishart, 1986), despite the attempts of some to downplay their musical relevance, claiming higher aesthetic value for “acousmatic” (see Schaeffer, 1966) or “spectromorphological” (Smalley, 1986) compositional praxes. The field of ecological acoustics, exemplified by research by such as Gaver (1993), Freed (1990), Warren and Verbrugge (1984) and Warren, Kim and Husney (1987) does offer empirical and theoretical support for auditory versions of Gibson’s notion of ‘lawfulness’ in visual perception (Gibson, 1966; 1979), and this paper will attempt to show how the direct and predictable connections between sound and source allow for a broader view of structure in electroacoustic music than that possible through developing the concept of ‘timbre’. Rather than attempt to capture the form of a music-analytical discussion of a some fragments from Yves Daoust’s intriguing (and self-parodic) piece Mi Bémol.

Sound as Information

According to an ecological approach to perception acoustic information, just like visual information, is a direct source of information about our surroundings (Gibson, 1966; Gaver, 1993). This directness is made possible by the predictable, or lawful, relationship between any sound and its causal event. For example, different styles of clapping produce different timbres because they involve different parts of the hand (Repp, 1987), and breaking and bouncing events may be distinguished by virtue of the lawful relationship between the temporal organisation of physical events (object impacts) and the sounds that they produce (Warren and Verbrugge, 1984). We become sensitive to such predictability through our exploration of a common and largely predictable environment (Gibson, 1966). Whether or not one wishes to accept the more extreme versions of the ecological approach, versions which deny the role of mental representation in perception, it is clear that the links between sounds and their sources are undeniable and, moreover, hard to disregard. The human auditory system seems attuned both to the need for segregation of the acoustic scene into separate streams and chunks (Bregman, 1990) and the need to identify and act upon the sources of these streams or chunks (Gaver, 1993). Moreover, given that such familiar events may be directly specified, it follows that any longer term sequence of sounds is liable to be perceived in terms of these events. As Howard and Ballas have shown, the ‘syntactic’ structure of a sequence of sounds interacts with the ‘semantic’ structure formed through the events that they may specify (see Howard and Ballas, 1980; 1982). Much electroacoustic music exploits this directness, and as this analysis aims to show, such directness does not reduce the ability of such music to create complex and intrinsic structural relationships. Rather than focus upon acoustic structure per se, therefore, it is instructive to focus upon the events this acoustic structure might specify, the environments they might occur within, and the interpretation of the relationships between different events and environments.
Real and Virtual Events
Before tackling the notion of environment it is useful to distinguish between two types of event, real and virtual. Both real and virtual events are specified in the same way: by a lawful relationship between acoustic structure and the physical interactions that produce that acoustic structure. In the latter case, however, perception is not veridical: the listener hears an event that has not occurred in the immediate environment (a recorded event), or an event which has never occurred (a synthesised event, such as that produced by Warren and Verbrugge, 1984).

The physical, real events which cause the sounds of a tape-piece to occur are the movements of loudspeaker components: although we can hear these as colouration of the sound it seems perverse to focus upon these aspects of sound rather than the virtual events which may be attributed to them by the listener.

Real and Virtual Environments
It is clear that the sounds we hear in Mi Bémol, regardless of whether they inform us about ‘natural’ events or more culturally specific events such as linguistic utterances or motivic structure, inform us about two kinds of environment. In the first instance the sounds of the piece serve as information for the real environment: although we hear sounds that have no visible sources it is clear that many more or less ambiguously defined events are specified by the acoustic structure of the piece. These events may not be perceived to occur in the listening environment itself, whether in a private space or a concert hall, but they can be clearly perceived to have their origins in a real environment as they preserve familiar and lawful relationships between events and acoustic structure. In the second instance, the sounds that we hear specify events within a virtual environment. In this case the sounds serve as information not for external events, but as the constituent elements of an intrinsic structure.

There are three main sources of information relevant to this distinction. First, it is clear that the listener’s relationship to these sounds is made within the context of ‘musical’ listening. The listener contextualises the sounds through the actions of putting on a CD, entering the concert hall and so on. This recontextualisation by the listener implies more than just a simplistic ‘aesthetic stance’ towards what is heard. The listener places the sounds within a context in which recorded sounds are going to be heard, sounds that by definition have already occurred, and which specify past events if they specify events at all. Second, and more importantly, many of the events do not normally occur within the kinds of environment which constitute a listening space, and hence cannot be considered to be lawful in relation to this space. Consider, for example the sounds of the outdoors which are to be heard in the piece, and the effects of a mismatch between the acoustics of the listening space and that of the recorded environment. The sounds of children playing (at 0'38” - 0'55”) do not just inform us about ‘children playing’ but about the acoustic space in which they play: the open air as opposed to the living room or concert hall. Third, the events specified by sounds in the piece do not follow an environmentally lawful pattern: the human listener remains stationary within a listening space, whilst events occur together which would imply many different environments: the outdoors; a firework display; the countryside. Sounds “fade in” in a way quite impossible in the ‘real’ world, and are recognisable yet clearly transformed or recombined in unlikely juxtapositions and superimpositions. Hence, the events specified through the acoustic structure of the piece can only be regarded as information for a past or historical environment, one which is in this case multiple and in itself unlawful in relation to the listening space and any single historical environment. In this sense the sounds of the piece present events which are part of a virtual environment which contains events which may be perceived to have occurred in the real environment or some possible future environment.

Intrinsic and Extrinsic Structure
Whilst listening to an acousmatic piece, events are more or less ambiguously specified, and relationships between these events are perceived. Those relationships which pertain to the real environment may be termed extrinsic, and those that pertain to the virtual environment, intrinsic. In the case of Mi Bémol some of the events are extremely familiar, if not banal, and some are extremely unfamiliar, to the extent that they can hardly be deemed to be “specified” at all.
First, some intrinsic relationships will be considered: those that seem to connect one section of the acoustic structure of the piece to another. Focusing at the shortest time scale, one should consider how it is that the smallest units of the piece come to be perceived as distinct and unitary. The opening sounds of the piece seem immediately to specify an instrument which, although unusual, is certainly coherent. The acoustic structure of these first sounds specifies a single event, a series of metallic, bell-like impacts, as if an object were being struck: the temporal structure of the series of impacts links them into a coherent, source-specific higher-order event. This temporal structure clearly specifies a decaying, bouncing form of event, similar to the synthesised bouncing event manipulated by Warren and Verbrugge (1984). This temporal coherence is important, since it serves to connect the impacts in an ecologically lawful manner. The first event suggests a single ‘instrument’ playing a short sequence of different pitches: in other words the acoustic information specifies that the original struck object is not only a single object, but a collection of objects of similar composition but different sizes. In this sense some form of instrument is suggested: the repeated striking of the same sequence of constituent objects does not suggest an inanimate set of objects being struck randomly (for example, a set of chimes being blown by the wind) but something that is intentionally ‘played’ by some agency: perhaps mechanical, perhaps a human player. The precision does not suggest the latter, but the melodic structure certainly hints that a human source is likely. The events which occur in the left portion of the virtual environment literally mirror those on the right, albeit with a time delay. The temporal relationships between these two sequence of events directly specifies two instruments, one on the listener’s right, and one on the left. The important point to note here is that although a musical interpretation, explicitly involving pitch and rhythmic structure could be found for these sounds, their coherence, distinctiveness one from another, and identity can all be explained in terms of the virtual events about which they inform the listener.

Consider again, for example, the events which can be heard between 0'38" - 0'55": in terms of a timbral description of this extract one might be tempted to state that a number of unrelated sounds had been superimposed. However, the cyclic metallic squeaking could easily be produced by the squeaking of a playground swing. It is not suggested that this is the only way to hear this sound. However, in this case such an interpretation provides a link with the sounds of children playing, despite the artificially foregrounded nature of the swing. The question of whether such a relationship is imposed upon these two sounds or specified will be returned to later. All that is required here is to suggest that as soon as one considers these sounds not as abstract timbres but as information for events, a shared origin in the real environment is made possible which implies a connection within the virtual environment. One should also note that the rather curious acoustic perspective, the ‘swing’ sound being louder and seeming artificially close to the listener compared to the actual ‘playground’ sounds, seems to simultaneously call this relationship into question. It should be added at this point that no claim is made that this is the only interpretation of these events. All that is intended here is to show how considering environmental specificity can draw a listener, in this case myself, into making some form of lawful interpretation of an artefactual and virtual environment.

There are many other intrinsic relationships this analysis could describe and discuss. Rather than continue in this vein, however, attention will now be paid to the relationship between the acoustic structures of the piece and the real environment of the listener. Firstly, a return to the ‘instrument’ events which open the piece is valuable in that it clarifies the ‘double articulation’ of acoustic information in acousmatic music (a double articulation hotly denied by Levi-Strauss, 1969). Just as the unitary nature of this ‘instrument’ forms the basis for defining a particular virtual environment, it simultaneously serves as information for the real environment of the listener. As already noted above, a degree of intentionality is specified by the acoustic information, offset by the mechanical precision of the ‘playing’, and also by the unusual and unfamiliar combination of bouncing impacts upon the ‘instrument’ and its similarity to some form of tuned percussion instrument. The real environment of the listener is directly specified by these sounds: the physical structure of objects and their interactions, and the familiar and more specific environment of musical instruments are both present within the relationship between listener, acoustic structure and the everyday environment. In more
general terms consider the following sources of information and their relationship to the everyday environment: a firework display, as discussed above, speech events (both children’s voices and news reports and interviews), outdoor events (animal sounds, weather sounds). All these sources of information may serve as information for the virtual environment of the piece, but simultaneously serve as information for a real environment, which although not immediately available, is lawful and persistent for the listener.

First of all, consider the superimposition of the pendular, metallic squeaking and the sounds of children’s voices mentioned a number of times above. Consider the last example again. Although it is stated above that the sounds of a swing and children playing are to be heard here, specific events, why is it that such an interpretation seems so reasonable? Although the structure of the sounds can be taken as specifying these events rather clearly, and their superimposition seems to strengthen their ‘association’ through a virtual space which although odd in terms of perspective, seems lawful enough, it is helpful to disentangle a number of factors here. Eco (1979, e.g. p. 190-192) suggests a concept that is useful in doing this: that of “motivation”. One can regard the structure of real and virtual environments not only as specifying events, but as shown above, as information that provides a context within which events are perceived. The real and virtual environments provide structures which “motivate” certain interpretations of events. Hence, as well as distinguishing between intrinsic and extrinsic relationships, one may also distinguish between intrinsic and extrinsic motivations which provide a context within which such relationships are perceived. In this case, the intrinsic relationship between the two acoustic structures is motivated by both intrinsic and extrinsic information, just as their extrinsic relationships respond to both forms of motivation. First, the sounds’ intrinsic relationship depends upon a pre-existing extrinsic relationship between swings and children’s play. Second, it relies upon the lawfulness of the real environment in which sounds that occur at the same time are more likely to be related than not, especially where visual information is unavailable, another extrinsic motivation. However, one can reverse this interpretation: the superimposition of the sounds within the virtual environment could be considered an intrinsic motivation which encourages a search for extrinsic relationships between the sounds and causal events, and between these events themselves.

Interpretation and Impoverished Information

In a certain sense, the acousmatic piece is an excellent example of impoverished information. The information we receive from a tape-piece cannot guide us around our immediate environment, nor does it inform us reliably and unequivocally about some historical or distant environment. Clearly, the structures it presents are impoverished in relation to specifying real environmental events and objects to which we must respond in a direct manner, events and objects which afford (Gibson, 1979). However it is not as impoverished as one might be led to expect by a traditional interpretation of the ‘acousmatic’. The information available from an acousmatic piece, as has become clear in the previous discussion of Mi Bémodal, is perceived within the context of the real and persisting environment, and is quite rich in its specification of extrinsic structures. More importantly, however, the acousmatic piece, however impoverished and unlawful it may be in relation to the real environment, is not necessarily impoverished in relation to its intrinsic structure, its virtual environment. Indeed, it is hard to talk of information being impoverished in a piece as rich in acoustic information as this.

There are two ways in which one can think of an acousmatic piece providing Gibsonian “affordances” (Gibson, 1979), perceptual structures which directly specify or constrain action. The first is that through its impoverishment and unlawfulness relative to the real environment it affords interpretation itself, a search for events and relationships between events which do provide for more immediate action on the part of the listener. However, this is supplemented by the second way in which the piece affords. The piece provides information which is perceived in relation to the real environment and the interpretative action of the listener: in this way the ‘meaning’ of the piece is not just to be found in its contrast to the real environment, but also in its similarities. The real and virtual environments involved in the perception of Mi Bémodal interact in such a way not only to provide information for interpreting the piece, but also information which interprets the environment. In this way, the piece may
itself be regarded as a cultural event which has an affordance structure. It affords multiply, but in a way constrained by its position within the environment. For example, the piece has afforded a discussion of its structure in this paper, just as it might afford some other form of analytical response. It could afford use as a cultural unit, as an exemplar of a particular style of music; it could afford communication of Amerindian land rights; it might afford contemplation on the part of the listener. Although these affordances are quite specific, it is not their perception itself which is of interest. Rather it is the way in which through listening to the piece the listener interprets not only the piece itself, but also, by virtue of this, the environment itself.

From the perspective of the piece, the environment provides the lawful background against which the listener’s perceptions and actions take place. From the perspective of the environment, the piece provides structures against which the environment is perceived and acted upon. It is not so much that the listener interprets the piece, rather that the piece, the listener and the environment may all be regarded as interpreting of the two other components in a triadic relationship, recalling Peirce (see e.g. Peirce, 1991):

```
piece

/\
/  \\
/    \
/

environment listener
```

In the analyses presented above, the listener was treated in a rather cavalier fashion, through presenting the kinds of explorations of real and virtual environment a single listener might make. Clearly though, the affordance resulting will be changed if any of these components is altered: differences in the environment (including social and cultural contingencies), in the piece, or in the listener’s interpretative actions will all create the possibility of a different affordance being perceived. Moreover, as noted above, it is difficult to clarify which of these mutually interacting sources of structure is responsible for a particular interpretative decision.

**Conclusions**

In conclusion then, one can view the perception of an acousmatic work as affording in two different ways: first, the mismatch between environment and piece relative to the listener may afford a search for information regardless of any observable change in the listener’s behaviour; second, the information gathered during this search may itself lead to the perception of new affordances. Covert interpretative action is no less real than the overt actions of booing and hissing or applause which greet new works in the concert hall. Such covert actions may lead to lasting changes in a listener’s behaviour which are difficult to trace in causal terms back to the experience of a piece. Indeed, in forming new interpretations (new cultural codes which might be shared and developed by a social body) the mutual relationship between the real and virtual environments involved in acousmatic music is no less pertinent to human ecology than any other change in our surroundings. The piece, then, is ‘about’ the environments upon which it draws, the environments which it portrays and our attempts to reconcile these two sources of structure, as much as it is about any abstract timbral, rhythmic or tonal scheme. Just as it would be a mistake to assume that a Mozart dance movement is a wholly abstract play of tonal and rhythmic forces, completely devoid of any connections to a broader cultural context, so too would it be mistaken to search for the significance of electroacoustic music through an examination of its inner workings alone.
References


391
Caractère musical de l’espace.

Nicholas Mamalis

(Crete University, Department of Philosophy, Section of Musicology and Theatrology).

L’élargissement des horizons sonores dans la musique de XXème siècle grâce à l’émancipation du timbre et ses relations complexes, la nouvelle technologie et la spatialisation de la musique, conduisent surtout vers une nouvelle problématique du matériau musical avec l’apparition des sciences cognitives de la musique.

Notre proposition a pour but de distinguer le caractère musical de l’espace parmi les autres dénominations (caractère du placement et de diffusion de l’œuvre) et de quelle façon l’organisation musicale nous conduit à la différence entre l’approche scientifique et la perception esthétique. Avec ce terme on essaie de préciser la dénomination musicale de l’espace de chaque œuvre qui veut fonctionner d’une manière esthétique et au-delà d’une utilité qui reste le plus souvent sa fonction.

Aujourd’hui, les différentes approches scientifiques du phénomène spatial étudient le son musical comme un multiple phénomène acoustique et sous trois disciplines différentes: a) Physique acoustique qui étudie la relation des ondes acoustiques avec les tympans des oreilles et les phénomènes acoustiques qui déterminent les qualités spécifiques b) La psychoacoustique qui étudie la relation entre les ondes acoustiques et la perception de l’imagination spatiale de l’auditeur et c) La neurophysiologie auditoire qui a pour champ d’étude la compréhension du système auditif, les structures neurologiques et les adaptations spéciales pour la localisation sonore1. Tant que les caractéristiques de propagation du son auront des qualités perceptibles (voir psychoacoustiques) donc affectives vont aussi avoir des qualités esthétiques.

Dans une problématique de dénomination musicale de l’espace, il faut d’abord noter sa fonction, puis les intentions d’utilisation du compositeur, ensuite les problèmes qui résultent de la manière à être perçu par l’auditeur et surtout l’importance de l’événement acoustique, qui dépasse son rôle d’un phénomène sonore de diffusion et d’amélioration de propagation de l’œuvre, et se valorise sous une fonction esthétique (voir musicale) étant perçu pour sa propre évolution.

Si l’on considère comme point de départ le déplacement d’un signal sonore d’un point de la salle à un autre, on n’accepte plus un tel phénomène comme un événement acoustique quotidien, mais on l’interprète aussi par rapport à la manière d’être traité sous relations avec les autres événements sonores de l’œuvre musicale. Sa relation comme un événement ayant une certaine signification réelle ne se limite pas en un simple rapport de déplacement. Notre conception change radicalement à partir du moment que l’on aperçoit un tel événement sonore dans une œuvre musicale. Sa

relation sémantique change considérablement en fonction des relations qui ont été établies avec les autres paramètres.

Sa fonction musicale se concentre sur l'événement lui-même et non pas sur l'événement acoustique comme diffusion sonore, apparaissant comme conséquence directe de son autonomie qui caractérise un phénomène esthétique. Ainsi la fonction de l'espace, quand il domine, se caractérise plus par rapport à son rôle esthétique qu'à son rôle de l'utilisation. Sa fonction esthétique devient toujours assez importante et elle ne peut pas participer sans influencer considérablement le résultat esthétique de l'œuvre musicale. Elle consiste à une partie du langage, refuse un simple rôle d'une fonction pratique et se transforme ainsi à un langage fonctionnel.

On peut mentionner que la dénomination musicale de l'espace, ne se définit pas par rapport à sa relation avec une réalisation d'un événement spatial vers laquelle il a été dirigé mais surtout par rapport à la manière d'être traité avec les autres composantes musicales. Si l'organisation de l'œuvre musicale montre une certaine hiérarchie temporelle des unités musicales (qui conduisent vers un langage et une syntaxe précise), un événement spatial se présente lié à un tel ou un autre moment de l'œuvre et apporte aussi son entité sémantique. C'est ce qui montre à l'auditeur que la signification que le compositeur a donné au phénomène spatial, peut être perçue par rapport à sa correlation avec les autres composantes.

Une autre constatation, qu'il faut aussi signaler est l'utilisation musicale de l'espace qui prend un caractère affectif. Un langage musical a assez de rapports communs avec un langage affectif et tous les deux ont une forte intention à montrer le sujet de l'énonciation. Et si dans un langage scientifique qui essaie de gérer le phénomène acoustique avec les différentes composantes et son but reste plutôt une détermination définitive des paramètres en question, la dénomination musicale, voir esthétique, donne plus d'importance à un certain moment du choix parmi d'autres qui sont possibles aussi.

Le rôle de l'invention scientifique (qualités acoustiques), reste plutôt de fixer les paramètres et les qualités de diffusion et de leur donner une substance objective en les définissant. Par contre, la domination musicale de l'espace donne plus d'importance au moment du choix, à l'intuition du compositeur et par conséquent il y a une volonté de devenir perceptible entre d'autres dénominations possibles.

Il a été souvent prouvé que les qualités qui caractérisent la musicalité de l'espace ont montré assez de désavantages comme l'efficacité qui a pour but la clarté, par ce que c'est possible que son utilisation ne conduise pas toujours vers une image auditive tout à fait nette. Un autre désavantage est souvent les limites des modèles proposés par ce qu'on peut trouver des manières de les traiter qui restent liées avec des techniques insuffisantes.

Ainsi on aboutit paradoxalement à la constatation que dans plusieurs œuvres dont les compositeurs ont utilisé l'espace comme paramètre de l'œuvre musicale, ce paramètre peut se faire valoir par rapport à son adaptation parfois hors du langage.
musical mais parfois aussi en relation avec son rôle qui lui appartient dans l'organisation de l'unité sémantique du matériau musical.  

Le caractère figuratif des aspects spatiaux peut rester indépendant du langage musical et se définir sous correlations arbitraires (fonction théâtrale, ornementale etc). Certaines œuvres, par contre, doivent leurs possibilités à la cohérence sémantique, cohérence qui permet l'introduction d'une fonction nouvelle et originale entre les qualités spatiales et la manière de les traiter. Et si sa fonction de diffusion tend vers un rôle d'une relation plutôt immédiate sa fonction musicale, elle, tend vers une relation totale.

---

Musical Development
Development of decentration in music listening in 3 to 8 year old children

Daisy Bertrand

Université de Liège, Unité de Recherche en Psychologie de la Musique, Belgium

INTRODUCTION

The development of children's ways of thinking in the field of music perception and cognition has been explored by different methodologies. For example, children were asked to identify two parameters presented simultaneously or to perform a well-known song with two pre-established constraints, i.e. to sing slowly but loudly (Sims, 1991), to detect chord changes when a melody is accompanied by chords (Costa-Giomi, 1994) or to associate movements to musical parameters and to move in relation to these when listening to the music (Sims, 1988). They were also presented two versions of the same tune and were asked to identify what was common and different (Zimmerman, 1984). These studies showed that young children from 2 to about 5 or 6 years old focus attention on only one aspect of the music among all the available ones and that children aged 7 and 8 years old were more able to direct attention to both parameters presented. These results are similar to those observed primarily by Piaget in non-musical tasks. However, these studies of the development of children's ways of thinking in the field of music perception generally involved two or three groups of children of contrasted ages and, to my knowledge, none of them focused on the transition phases between the different ways of thinking.

The present study focuses on the development of decentration of attention in music listening in children aged 3 to 8. It aims to evaluate whether the children’s ways of thinking change progressively or suddenly. Visual representations of the musical parameters were used and the children had to point out the parameters present in the music they were listening to.

METHODOLOGY

Eight musical parameters were used to construct 148 sequences of 9 tones. These parameters were timbre, density (number of simultaneous tones), pitch, intensity, articulation, length of the tones, introduction of a silence and introduction of a longer tone amongst shorter ones. In each sequence, only two parameters were changed while the others remained constant.

Each musical parameter was symbolized using a domino made up of two pictures representing both values of the parameter. For example, low intensity was symbolized by a small drum and high intensity by a large drum. The articulation parameter was depicted by a walking boy representing the slurred tones and a skipping girl symbolizing the dotted tones. The same kinds of pictures were used to represent the six other parameters. The pictures
which the children considered best depicted the musical parameters presented were selected on the basis of a pre-test.

Before undergoing the testing sessions, children first underwent training sessions using different tasks in order to learn the relationships between the parameters used in the sequences and their pictorial representations. However, only children older than 4 were able to detect the two temporal parameters (the introduction of a silence or the introduction of a longer tone amongst shorter ones) without assistance. Younger children were able to respond according to these parameters only when the experimenter was helping them by exaggerating the length of the silence or the duration of the lengthened tone. Without this help, they were unable either to recognize the parameter or to locate correctly the rest or the lengthened tone in the sequence. Therefore, the sequences containing these parameters were not presented to children aged 3 and 4.

During testing, the dominoes relating to the parameters used in the sequence were placed in front of the child. While listening to the music, the child was asked to point out all the parameters present. Each sequence was presented two times. The sequences were run in 5 to 10 sessions depending on the child, as the session was stopped when the child became less attentive.

34 children from the three grades of a nursery school (children from 3 to 5 years old) and the two first grades of a primary school (children aged 6 to 8) were included in this experiment. None of them had received any musical training nor had participated in musical activities. All these children successfully learned the relationships between the parameters and their pictorial representations during the training sessions.

RESULTS

As the children were asked to point out all the parameters present in the sequence, they should have identified 4 pictures for each sequence. However, the 9504 responses collected were very different from each other and could be gathered into 34 different types.

1. A psychometric analysis of these responses was first achieved using the SAS Prinqual procedure (SAS Institute Inc, 1990). This data transformation procedure is similar to a correspondence analysis but assumes that the categories of a nominal variable correspond to values or scores of a single interval variable. This one-dimensional representation of response type was appropriate as reliability is about 0.90. When each response type is ranked by its score, it is apparent that the one-dimensional representation resulting from the prinqual procedure can be considered as a centration-decentration scale. According to these scores, children's responses can be gathered roughly into three main groups indicating the characteristics of the children's way of thinking.
Low scores were found when responses were characterised by centration during the entire sequence (when children pointed out one parameter at a time during the whole sequence). Intermediate scores were observed when children were able to decentrate during a part of the sequence, either during the beginning of the sequence or after the parameter change. In these cases, two parameters are identified simultaneously during a part of the sequence and only one during the rest of the sequence. Finally, high scores were given to responses reflecting children’s ability to decentrate attention during the whole sequence. In this case, two parameters were identified at a time during the whole sequence.

2. A statistical analysis was used relating the scores obtained by the Prinqual procedure to the characteristics of the child. An anova analysis of the scores attributed to the responses with 2 nested factors (age, subject (age)) reveals that the scores of children’s responses were strongly influenced by age ($F(5,3792) = 3359; p<0.0001$).

As shown in figure 1, responses given by the 3 and 4 year old children receive the lowest scores, and the scores increase with age. However, the increase of children's scores is not linear. No significant rise is observed between the age of 3 and 4 and between the age of 7 and 8. Scores increase significantly between age 4 and 5 and between age 6 and 7 but the main evolution occurs between the age of 5 and 6.

A qualitative analysis of children’s responses by age is presented in figure 2. Responses given by the 3 year old children reflect their centration of attention on one musical parameter at a time. This is also observed for the 4 year old children, although their responses reflecting centration are slightly lower and those reflecting decentration during the end of the sequence.
are slightly higher. An evolution is observed in the responses of the 5 year old children. The ability to decentrate seems to emerge around 5 as responses reflecting decentration of attention become more frequent and are observed either during a part of the sequence or during the whole sequence. However, responses reflecting centration of attention are still common. A major change in the development is observed between the age of 5 and 6, as the 6 year olds' responses are strongly characterised by decentration of attention during the whole sequence, although centration and partial decentration responses are still present. A further evolution is observed between the age of 6 and 7 as nearly all the 7 and 8 year old children's responses are strongly characterised by decentration of attention, and only a few centration and partial decentration responses are observed.

3. A third analysis focused on the variability of children's responses in the same group of age. An anova analysis of the scores attributed to the responses with 2 nested factors (age, subject (age)) reveals a variability between subjects nested within age (F(28,3792) = 134.56; p<0.0001) (Figure 1 and 3). The inter-individual variability is rather low at the age of 3 and 4 and at the age of 7 and 8. This means that children of the same age give the same kind of responses. On the contrary, the type of responses varies highly from one child to another in the 5 and 6 year old group as variability is significantly high. A further investigation of the 5 years old's responses shows that nearly all of the 5 year old children's responses were given lows scores (figure 3). These children presented the same response's profile and gave many responses characterised by centration during the whole sequence. One received an intermediate score for his responses, reflecting some ability to decentrate attention during the whole sequence. The last child presents a completely different profile to the other children in this group. His responses are mainly characterised by decentration during the whole sequence. However, centration and 'partial' decentration responses are still present. None of the 6 year
old children received low scores for his responses and only two of them received intermediate scores resulting from the high number of responses characterised by centration of attention. The other children's responses are mainly characterised by decentration and centration responses are rather infrequent.

![Figure 3: Scores and standard deviation by subject](image)

4. The last analysis focused on intra-individual variability (figure 3). Intra-individual variability is rather low for all the 3 and 4 year old children and the 7 and 8 year old children. This means that a child gives the same kind of responses during the whole test. On the other hand, important intra-individual variability is observed for some of the 5 year old children and most of the 6 years old. This means that a 5 or 6 year old child gives many different types of responses to the 148 sequences of the test. It is apparent that high variability is only observed when an intermediate or a rather high score is given to the child's responses and that variability is low when its responses were given a low or a high score. This suggests that high variability is observed when the child is moving from one way of thinking to another, either immediately preceding or following this transition.

CONCLUSION

The present study shows clearly that the 3 and 4 year old children’s ways of thinking is highly characterized by centration of attention. This means that their attention is mainly focused on one musical parameter to the exclusion of the others and that they are only be able to take into account one musical parameter at a time. Children's ways of thinking seem to undergo a slight change between 4 and 5 years of age, as some children are more able to decentrate attention for a small period of time or even during the entire sequence. However, most children's way of thinking is still highly characterised by centration. The major change
in the development of children's ways of thinking occurs between the age of 5 and 6. Nearly all the children are now able to direct their attention to more than one musical parameter at a time, and this ability grows with age until the age of 7 or 8. These children's ways of thinking is then highly characterised by decentration. These findings are in accordance with the results observed in previous studies (Sims, 1988, 1991, Costa-Giomi, 1994, Zimmerman, 1984) as they observed that the transition from one way of thinking to another occurs between 5 and 7 years old.

Because the present study involved children of the full range of age between 3 to 8 rather than only two or three contrasted groups of age, the nature of the development of children's ways of thinking could be observed. It is apparent that this development is not linear. On the contrary, it starts slowly, speeds up with age and slows down after the age of 6.

The changes involved in this development are not quantitative but consist of a sudden qualitative difference. Actually, the changes observed did not consist in a increase or a decrease of a type of responses but in a change of the type of responses given by the children. This qualitative change reflects a reorganisation of cognitive structures as the ways of thinking change from simple, centration, to more complex, decentration.

It appears also that this development is not progressive as no intermediate states is observed between the two way of thinking, centration and decentration. Children's scores are distributed in a bimodal way on the centration-decentration scale. Their scores are either low or either high and only very few children received an intermediate score for their responses. This means that their responses are characterised either by centration, either by decentration during the whole sequence, and that decentration during a part of the sequence is rather rare in children's responses.

Finally the high inter-individual and intra-individual variability observed at the age of 5 and 6 reflects the loss of stability that occurs close to the transition phase. The nature of the differences between the children changes as they are ready to make the jump to decentration.

REFERENCES
Quantitative and Qualitative Approaches to Understanding Children’s Perception of Musical Pitch in Development

Alexandra Lamont
University of Cambridge, School of Education and Gonville & Caius College

This paper aims to develop our understanding of children’s representations of pitch relationships in Western tonal music by combining quantitative and qualitative approaches. The quantitative approach focuses on the well-established probe-tone technique as a means of uncovering pitch representations within a short-term listening situation, whilst in the qualitative approach children engage in a range of different interactive musical tasks intended to pose perceptual difficulties and thereby to illuminate their representations of pitch.

Experimental research with musically trained adults has tended to provide support for the perceptual reality of the organisation of pitch in tonal music according to the principles of music theory, showing a relatively rigid hierarchy of notes, chords and keys to be a powerful perceptual organising force (e.g. Krumhansl & Shepard, 1979). The abilities of ‘non-musicians’ have been less examined and consequently are less well understood, but the general conclusion is that they are much less sophisticated than those of ‘musicians’. However, this broad category of ‘non-musician’ requires further investigation as it is likely that it obscures considerable diversity in terms of prior musical experience, and it also includes a greater section of the population. Our understanding of adults’ pitch perception may be considerably extended through examining how these develop over time and attempting to establish the factors which may affect or determine this in different ways. Previous developmental studies focus on certain critical ages in the development of pitch processing: for example, infancy (Trehub & Trainor, 1993; Lynch et al., 1991) or children aged between 5 and 11 (Krumhansl & Keil, 1982; Bartlett & Dowling, 1980). Despite the significant differences between children aged 11 and adults, the period between 11 and adulthood has been very little studied. In addition, developmental studies rarely differentiate between ‘musician’ and ‘non-musician’, considering children simply as naive listeners. This means that we know little about how training can influence musical development, either in adults or in children.

The current study attempts to address some of these issues through a focus on children aged between 5 and 15 and through the inclusion of different levels of musical experience. Musical experience can be defined using Sloboda’s (1985) distinction between enculturation, referring to the implicit understanding of one’s musical culture acquired as part of the complex process of development, and training, referring to the explicit detailed and rigorous programme of instruction given to those wishing to pursue music professionally. Hence general music education in Britain aids the process of enculturation, introducing children to the materials of music and the possible ways of combining them, from a listening and a performing/composing perspective. Entrance to the musical élite requires a further specialist training programme, which often focuses around the acquisition of performance skills on a particular instrument. There are evidently many stages between the lowest degree of educational enculturation provided by class singing or music listening and the highest degree of training provided by specialist tuition, and thus any investigation of musical abilities in development must account for musical experience in a more sophisticated manner than labelling children as either ‘musicians’ or ‘non-musicians’. This study attempts to address experience-related development through the inclusion of as many different levels of musical experience as possible across the age-range.

Firstly the results of a large-scale listening task using the probe-note technique with a range of children aged between 5 and 15 will be presented. The melodic contexts consist of an arpeggiated tonic triad (C-E-C-G in C major), ascending complete major scale (C-D-E-F-G-A-B-C), and arpeggiated diminished triad (B-D-B-F), as previously used in studies with both children and adults by Cuddy & Badertscher (1987) and Brown, Butler & Jones (1994). The first two contexts present tonally well-structured sequences, and these evoke different response profiles from children who have undergone formal training and children who have not. From the major triad context, formally trained children show a clear distinction between diatonic and
non-diatonic notes, with further preferences for the mediant E, tonic C, dominant G and subdominant F, whilst untrained children show a very similar pattern but with significantly less differentiation. Both groups give very low ratings to the leading note B below the initial tonic, which is rated equivalently to the non-diatonic notes. From the major scale context, formally trained children show again a clearly defined profile with highest ratings for the tonic C, followed by the mediant E and then the other diatonic notes. This context produces response profiles for untrained children which clearly show the influence of pitch height. As the diatonic scale ascends, so notes at the top end of the range are strongly preferred over notes separated by a larger pitch interval, with less distinction between diatonic and non-diatonic notes (as found for musically untrained adults by Krumhansl & Shepard, 1979). These rating profiles are comparable to the overall findings of the earlier studies.

The diminished triad context is more problematic in nature, as it does not explicitly present significant elements of the traditional tonal hierarchy established by Krumhansl et al. (see Krumhansl, 1990), yet the rare interval hypothesis proposed by Brown et al. (1994) suggests that the outlined tritone-plus-one figure should unambiguously determine identification of the correct tonal centre. Brown et al.'s empirical findings with adults seem to suggest that this is in fact the case, showing higher ratings for the tonic and notes of the tonic triad than any other probe notes, although Cuddy & Badertscher (1987) were only able to establish the tonal hierarchy in two highly-trained adult subjects when these subjects were asked to respond 'theoretically' rather than figuratively to this context melody. The current results suggest that this context is problematic for children also, as different profiles were found from children with three different levels of musical experience. These three levels were also significant in an earlier study using randomised diatonic contexts in a probe-note methodology (Lamont, 1996). Children with no musical experience whatsoever produce a relatively flat profile with small peaks at the supertonic D (presented once in the context) and mediant E. Children with practical involvement with music show similarly flat profiles but preferring most strongly the supertonic D and a small peak at the leading note B (presented twice in the context), suggesting an increased sensitivity to the properties of the context. Children with formal training produce far more defined profiles, with highest ratings for the supertonic D, sharpened subdominant F# and leading note B, and with a differentiation between other notes in a direction contradictory to the expected tonal hierarchy. Their ratings for B and D are explicable in terms of event hierarchy information, namely as reflecting the properties of the context melodies, whilst the high rating for F# suggests an outwards rather than inwards resolution of the tritone interval.

Figure 1 shows the extremes of these differences in a comparison of profiles from younger non-musicians and older trained musicians.
These results directly conflict with those of the earlier studies (which contradict each other also). They do however suggest that a method using simplified musical materials is an appropriate way of investigating pitch perception only if carefully implemented, as both children and adults appear to be very sensitive to the properties of musical materials which are presented to them in such short-term listening tasks.

These findings will be compared with children’s performances on a range of interactive musical tasks which can shed more light on these differences and difficulties in perception. The tasks use chime bars, which are frequently used in school musical activities and hence familiar, as a way of embodying pitch in tangible objects that can be manipulated by the children. A set of chime bars was customised for this study similarly to the Montessori bells used in a similar context by Bamberger (1991): a full chromatic set was painted blue and a diatonic set painted red, to remove the clues of white and black notes from the original materials. Sessions were conducted with children on an individual basis in the manner of a series of games (following Lamont & Cross, 1994), and were video-recorded.

The initial tasks also draw inspiration from Bamberger’s study: the children begin with a pitch-matching task which evaluates their ability to recognise chime bars which make the same sound, and continue with a tune-building exercise using the familiar children’s song of ‘Twinkle Twinkle Little Star’. This song, described as a ‘structural simple’ (Bamberger, 1991: 182), consists of two complementary phrases on the pitches C-G-A-G F-E-D-C as follows:

```
C-G-A-G F-E-D-C
```

Twin- kle twin- kle lit- tle star How I won- der what you are

After successful building of this tune from a chime bar set where each crotchet note is represented by a single bar (C-G-A-G F-E-D-C), the second bars for the duplicated notes G and C are removed and the children are again asked to try to perform the tune: a perceptual problem is thus posed. Their ways of solving this problem illuminate their underlying representations of pitch: children with formal representations will rapidly find the solution as they are able to deconstruct the phrase into its constituent elements (the ability to focus on the time-independent aspects of pitch), whilst children with figural representations have more difficulty as they are unable to deconstruct the figure into its parts (their representation of the figure remains as a gestalt).

On the basis of a number of aspects of the children’s performance, a classification into formal and figural representations is proposed. This includes a consideration of the initial tune-building strategies, ways of solving the problem set, time taken, the kinds of assistance required, and the children’s confidence. This classification is far from straightforward, with many children exhibiting both types of representation at different points during the task. It is only those children who are experiencing formal musical training, however, who demonstrate a consistently ‘formal’ perception of pitch. Although these children are also often older, purely figural representations are observed in older children who have not experienced formal musical training. This provides some support for a Brunerian phase-based model of development such as that outlined in a musical context by Bamberger (1991), where experience creates differences in perception which may not be achieved by simple maturation and general enculturation.

The subsequent task brings the listening probe-note methodology to life. Context melodies are played to the children using the red diatonic bars, and they sort the blue chromatic bars into groups according to their goodness of fit with the melodies. On a first pass the children are asked to make bipolar decisions between ‘good’ and ‘bad’, and a second stage allows for more sophisticated choices amongst the resultant groups of ‘good’ and ‘bad’. These sessions provide very detailed information about the children’s decisions and their verbalisations of the process.
The following discussion is based only on a sample of results from children aged between 6 and 11 for the two contexts of major triad and major scale (data will be presented for a greater age range of children and additionally for the diminished triad context). One striking feature from this note rating task is the consistency between note ratings from the tonic triad and scalar contexts, suggesting (contrary to observation) that in fact children's judgements on this kind of task are highly consistent. By about age 8, most children appear to prefer notes prominent in music theory, with a consistent overall preference for the note C4 from both major triad and major scale contexts. However, non-diatonic notes are often found amongst those considered 'good', and certain diatonic notes may appear in the 'bad' groups. These results are also not easily explicable according to psychoacoustic principles, with pitch height playing only a marginal role in the scale context for some younger children, although the children's own verbal descriptions of their judgements frequently draw upon pitch height as a basis for judgement.

The differences between children with 'figural' and 'formal' perceptions of pitch are again not readily apparent, and one reason for this may be that the 'figural' perceptions of the context melodies can differ. For example, from the scalar context, certain children explain that they like 'high' notes at the end, because the scale rises, whilst others describe the need for the figure to fall in pitch at the end, more concordant with the notion of gap-fill melodic structure. Nonetheless, the musically trained children approach the task in a different way and one which may also be subject to age-related change. Some older trained children show a tendency to reject some of the notes which are present in the context, especially for the tonic triad context type, considering them to be insufficiently interesting. Furthermore, these older trained children are also able to make more sophisticated judgements than the two-value task which was posed for them, often placing a group of notes in order of preference or asking for new categories to be created for their judgements. Analysis of all the results will be presented to provide a detailed underpinning for interpretation of the results of the larger-scale listening study, and hence to further our understanding of children's perception of musical pitch.

References


Cuddy, LL & Badertscher, BD (1987) 'Recovery of the tonal hierarchy: Some comparisons across age and levels of musical experience' *Perception & Psychophysics, 41*, 609-620


Krumhansl, CL & Keil, FC (1982) 'Acquisition of the hierarchy of tonal functions in music' *Memory & Cognition, 10*, 243-251

Krumhansl, CL & Shepard, RN (1979) 'Quantification of the Hierarchy of Tonal Functions within a Diatonic Context' *Journal of Experimental Psychology: Human Perception & Performance, 5*, 579-594


The Development of Pitch Relations in Adulthood: Effects of Training and Education

Daisy BERTRAND *, Alexandra LAMONT**, Ian CROSS***, & Irène DELIEGE*

* U.R.P.M., C.R.F.M.W, Université de Liège
** School of Education & Gonville & Caius College, University of Cambridge
*** Faculty of Music, University of Cambridge

This research evaluates the effects of general education and specialist music training upon the degrees of sensitivity that adults have to the diatonic and tonal organising structures of Western tonal music. The study employs a short-term listening task following the probe-tone technique developed by Krumhansl (1979), where musical sequences (such as the diatonic scale or a cadential sequence) are presented as contexts followed by the twelve chromatic notes as probes to be rated for goodness of fit with the preceding contexts.

When considering the differences between musically trained and untrained listeners, the issue of context types appears to be important. With highly-structured context materials, different response strategies are observed from musically trained and musically inexperienced adult subjects. For example, Krumhansl & Shepard (1979) found a progressive differentiation within rating profiles with degree of musical experience. The most highly experienced group produced responses consonant with music theory, preferring strongly the tonic, tonic triad and diatonic scale amongst the twelve chromatic notes. The moderately experienced group preferred the tonic and diatonic scale only, whilst the least experienced group showed only a preference for the tonic alone. The responses of this least experienced group were considered to be “non-musical”, appearing to be based upon pitch height rather than tonal relationships. Results from studies by Krumhansl and her colleagues can thus be read as implying that musically trained subjects have an advantage over non-musicians in their capacity to access any knowledge that they have acquired of the hierarchy of notes within the tonal system. Furthermore, when contexts are deliberately structured to contradict listeners’ tonal expectations (as in a study by Oram & Cuddy, 1995), musicians again produce profiles strongly influenced by a long-term tonal hierarchy, whether present or not (here, additionally influenced by frequency of occurrence of notes within the contexts). Non-musicians showed no influence of the tonal hierarchy in making their judgements, their responses being determined primarily by pitch proximity.

However, when randomised diatonic context sequences were employed - thus minimising the cues to tonal function afforded by the temporal characteristics of the context sequences (such as consistent presentation of particular patterns of notes in specific orders, or use of different frequencies of occurrence or different total sounding durations for different notes within the sequences) - musically trained subjects respond similarly to non-musicians (West & Fryer, 1990). This suggests that this modification of the probe-note technique succeeds in removing the advantages conferred on musically trained subjects through the use of structured materials, and thus results may be more reliably interpreted as indicative of short-term musical processing.

In the above research, the defining criterion for categorising subjects as musicians or non-musicians is extent of instruction in practical music and/or formal training in music theory. Musicians in these studies tend to have more than five years of training of some kind, whilst non-musicians tend to be those subjects who do not learn musical instruments and who report no formal musical training. However, results from a developmental study (Lamont, Bertrand, Cross & Deliège, submitted) indicate that this bipolar distinction constitutes too impoverished a framework to adequately address the varieties of musical experience that may impact upon listeners’ judgements. Between the ages of 6 and 11, the presence or absence and amount of both general music education and formal musical training are seen to have a clear effect upon representations of pitch organisation. Thus the question of musical experience in adulthood,
and the potentially lasting effects of both types of musical experience, deserves more detailed enquiry.

**Method**

Context sequences were produced by randomising the seven notes of the diatonic scale, and followed by 12 probe notes. Two different randomisations were produced for each probe note in two different conditions: final intervals ascending or descending, producing a total of 48 sequences which were randomly ordered and transposed to different keys for presentation. Over three hundred adult subjects from Belgium and England participated in the experiment. They were asked to rate probe notes for goodness of fit with the preceding context melody on a five-point scale.

Information was collected on subjects' musical backgrounds in terms of current and past musical training, participation in musical activities, and so forth. We find two groups of 'non-musicians', one (Belgians) with no formal musical training nor any general music education, and the second (English) with no formal training but having experienced general music education in school. At the other end of the scale, there is a group of 'expert musicians' (English), who have undergone a substantial amount of formal musical training and practical involvement (between 7 and 18 years of training) as well as general music education in the past. Between these extremes, there lies a range of different musical experiences including subjects with musical training in the past (both English and Belgian) as well as adult subjects who begin formal training in music in adulthood (only Belgian). We will consider analysis of rating profiles for each of these groups in turn.

**Analysis**

**Non-Musicians**

The Belgian non-musicians (N=48) have a mean age of 23 whilst the English non-musicians (N=21) have a mean age of 30.5, both groups comprising university students. These two groups of non-musicians do not give significantly different responses to the materials, suggesting that the presence or absence of general music education in the past does not create any differences in response strategies. The non-musicians' profile is shown below.

![Rating Profile for Non-Musicians](image)

This profile for non-musicians shows a significant difference amongst probe notes ($F(11,7370)=19.057, p<.0001$), with a clear differentiation between diatonic and non-diatonic
notes (F(1,737)=92.894, p<.0001). There are no other obvious characterising features of this profile (no especially high ratings for the tonic or notes of the tonic triad, for example).

Final Interval Direction is found to be significant (F(1,67)=6.473, p=.0133), sequences with ascending final intervals being given higher ratings than sequences with descending. There is also a significant interaction between the two variables of Probe Note Identity and Final Interval Direction (F(11,737)=7.936, p<.0001). Both the ascending and descending profiles analysed separately show significant differences between ratings for diatonic and non-diatonic notes (planned contrasts from ascending sequences (F(1,737)=68.317, p<.0001) and descending sequences (F(1,737)=29.460, p<.0001) are both significant). However, the profile for descending sequences is less differentiated. For the four notes which emerge as highly preferred in the global profile for this group, two are most significant in the ascending condition (E and F#) whilst the other two are most significant in the descending condition (F and G).

**Expert Musicians**

This group comprises English university students studying for degrees in music (N=77) with a mean age of 20. They have a mean amount of formal musical training of 13 years, and the mean age of starting this training is 6.5.

No significant interaction is found between ratings for Probe Notes and number of years of musical training (F(11,704)=.93, p=.6849), nor between ratings for Probe Notes and age of starting musical training (F(11,704)=.95, p=.6249). This suggests that after a certain amount of training, subjects' representations of pitch stabilise and the pitch schema they have developed is reliably applied to new listening situations.

There is a significant main effect of Probe Note Identity (F(11,836)=129.34, p<.0001). The overall ratings show a significant differentiation between diatonic and non-diatonic notes (planned contrast shows F(1,836)=1225.615, p<.0001), and a further distinction between the notes of the tonic triad (C, E and G) and the other diatonic notes (F(1,836)=120.589, p<.0001). In addition, ratings for C, F and G are higher than for other diatonic notes (F(1,836)=48.22, p<.0001), and the tonic is rated significantly higher than all other diatonic notes (F(1,836)=51.753, p<.0001) as well as above the other two triad notes (F(1,836)=5.233, p=.0224). Amongst the non-diatonic notes, the rating for F# is found to be significantly higher than those for other non-diatonic notes (F(1,836)=41.34, p<.0001).

---

**Rating Profile for Expert Musicians**

![Rating Profile for Expert Musicians](image)
Again, sequences with ascending final intervals received higher ratings than sequences with descending final intervals \( (F(1,76)=33.038, p<.0001) \). The interaction between probe note identity and final interval direction is also significant \( (F(11,836)=8.171, p<.0001) \). Both profiles analysed separately show a clear differentiation between diatonic and non-diatonic notes and within the diatonic notes a clear preference for the notes of the tonic triad and the tonic itself. However in the descending condition this pattern of differentiation within the diatonic notes is more clearly marked.

**Middle Group of Musicians**

We can initially consider there to be three main groups between the two extremes of non-musicians and expert musicians. The first comprises Belgian ‘past musicians’ (musical training experienced in the past, \( N=43 \)), with a mean age of 22.2, mean amount of 5.5 years of training and a mean starting age of 10. The second is English ‘past musicians’ (\( N=44 \)), with a mean age of 25, mean amount of 6.4 years of training, and mean starting age of 9.3 years, and the final group is. Belgian ‘current musicians’ (musical training begun in adulthood, \( N=75 \)), with a mean age of 33, mean amount of training 4.2 years, and mean starting age of 24.

However, probe note ratings do not differ significantly between these three groups. Firstly, there is no significant difference between Belgian and English ‘past musicians’. This indicates that the general music education which the English subjects underwent during childhood does not give them any advantage (or disadvantage) by adulthood over Belgian subjects whose musical development can be ascribed only to general enculturation (as for non-musicians). Furthermore, there is no significant difference between these ‘past musicians’ and the Belgian current musicians. The overall profile is shown below.

However, one differentiating variable amongst this group is the number of years of training, which significantly interacts with Probe Note identity ratings \( (F(154,1445)=1.329, p=.0063) \). The following analysis compares only Belgian past and current musicians, as the training programme they have experienced is highly structured and comparable. In general, analysis of these profiles for each successive year of training shows that profiles from the ‘past training’ subjects appear to be better defined than those from the ‘current training’ subjects at the same level of training. The mean starting age for the ‘past training’ subjects is 10, whilst the mean for the ‘current training’ subjects is 24. This suggests that the formal training
programme offered in Belgium has more effect upon subjects’ development of pitch relations when it is experienced during childhood than during adulthood.

Conclusions

Although we have shown elsewhere that general music education has an effect upon children’s responses to musical pitch, the results of the current study suggest that this effect does not persist in differentiating between groups of adults. It seems probable that adults’ experience with and exposure to music is in itself sufficient to level out any differences created by earlier general music education. The profiles observed from the non-musicians in this study can be interpreted as a levelling up rather than down, as they do show a sensitivity to tonal relations that has not always been found in earlier research and that can only be ascribed to enculturative factors.

The difference observed here between adults who had experienced some degree of musical training in the past and those who had not is an important finding, and implies that even if musical training had been experienced a considerable time in the past, this factor is responsible for the more sophisticated internal representation of pitch within the tonal system shown by the ‘past musicians’ in this study. However, the lack of difference in profiles between English and Belgian ‘past musicians’ is also interesting, as the formal training programmes which these subjects experience is very different; furthermore, during childhood we have found significant differences between these two groups. The different schemata developed as a result of different types of formal training in childhood appear to resemble one another after time, which may indicate a stabilisation process that results from the evocation of these schemata in musical perception over a number of years.

The profiles of Belgian ‘past’ musicians, when compared with those of Belgian ‘current musicians’, appear also to be more stable and show greater differentiation amongst notes. This may again be due to the greater number of years these subjects have been able to apply their pitch schemata, although additionally this points to the greater effectiveness of learning in childhood.

Finally, the amount of training appears to be a very important factor in explaining the degrees of sophistication observed in rating profiles. This is especially clear for the Belgian musicians, which reflects the structured nature of the training programme in Belgium. The English ‘expert musicians’ produce a very clearly defined and differentiated profile with a significant amount of musical training, and results from this group suggest that after a certain threshold the actual amount of training ceases to produce any further qualitative change in responses.

References


The research reported here was funded by an award from the British Council, the Fonds National de la Recherche Scientifique, and the Commissariat Général aux Relations Internationales. Daisy Bertrand was supported by the Fonds National de la Recherche Scientifique, and Alexandra Lamont by the Economic and Social Research Council.
Effects of the pitch relationship between text and melody in Cantonese songs on young children’s singing

Lily Chen-Hafteck
ASME, Roehampton Institute London, U.K.
Dept of Music, University of Pretoria, South Africa.

Introduction

In the literature of children’s singing research, the effects of the text-melody relationship has not yet been clarified. Some researchers argued that words are a distraction to children’s song-learning and the learning of melody before words can possibly help improving pitch accuracy in children’s singing (Goetze, 1986; Levinowitz, 1989; Welch, Sergeant & White, 1995/96). However, the results of this research are inconsistent, with numerous studies failing to show this effect (Sims, Moore & Kuhn, 1982; Levinowitz, 1987; Smale, 1988). Furthermore, the research on adult and children’s memory of songs (Serafine, Crowder & Repp, 1984; Serafine, Davidson, Crowder & Repp, 1986; Crowder, Serafine & Repp, 1990; Morrongiello & Roes, 1990) has suggested that text and melody are integrated in one’s cognitive processes. This further confirms that text-melody relationship is an important issue in music education. However, the role played by the text-melody relationship in children’s song-learning and singing still remains uncertain. Whether the degree of text-melody relationship exerts effects on children’s singing is a research question that needs further investigation. The present study examined this issue with reference to Cantonese children’s songs.

Cantonese is the Chinese dialect spoken in Hong Kong. Due to the tonal nature of this language, each Cantonese word has its relative pitch inflection, which has to be respected in melodies of songs so that the sung words will sound most meaningful. However, the Cantonese children’s songs commonly sung in Hong Kong are dominated by Western nursery rhymes, with translated Cantonese texts disregarding the relationship between melodic contours and linguistic tones. Whether the discrepancies between texts and melodies in the current Cantonese children’s songs affect children’s song-learning and singing and whether Cantonese words can be fitted into Western melodies are the main concern of this research.

Objectives and hypotheses

The aim of this study was to clarify the effects of the pitch relationship between text and melody in Cantonese songs on young children’s song-learning and singing. It was hypothesized that songs with tone-mismatched melodies (where the linguistic tones in the text does not match with the melodic contour) will be performed with lower accuracy than songs with tone-matched melodies (where the two match).

Another objective was to address the issue of fitting Cantonese texts to diatonic melodies. It was assumed that Western diatonic melodies will be more closely matched
Effects of the pitch relationship in Cantonese songs
to English texts than Cantonese and thus, a diatonic song with an English text will be performed with higher pitch accuracy than a Cantonese text. Moreover, since English and Cantonese languages have different pitch inflections, certain intervals may be sung more accurately to an English text and others sung more accurately to a Cantonese text. This study attempted to identify these intervals.

It was also predicted that the effects of text-melody relationship on children’s singing are not limited only to pitch accuracy. Therefore, other aspects of singing including the singing style, text recall and song-learning processes, were also studied.

Methodology

Experiment One

Twelve children (10 girls; 2 boys) in Hong Kong aged between three and five who spoke Cantonese as their mother-tongue were the subjects of this study. These children were taught three Cantonese songs. After one month, their singing of these songs was recorded. No starting note was provided.

Song A: a tone-mismatched song (where the pattern of the linguistic tones in the text does not match with the melodic contour);
Song B: a tone-matched song (where the pattern of the linguistic tones in the text match the melodic contour);
Song C: a traditional Cantonese nursery rhyme (a chant to be recited without fixed melody).

Pitch analysis was carried out by listening to the recordings and transcribing them into score notations. Intervallic relationships and melodic direction were chosen as the criteria in assessing pitch accuracy in order to clarify whether the tonal relationships of texts have any effect on the degree and direction of pitch movements that children sing. The quantitative analyses performed included:

(a) Accuracy in intervallic relationships of Songs A and B
(b) Accuracy in melodic direction of Songs A and B
(c) Pitch accuracy of individual intervals of Songs A and B
(d) Average intervals performed in Song C
(e) Accuracy in pitch inflections of Cantonese tones in Song C

Beside quantitative analysis of pitch accuracy, some qualitative considerations were also included. Some descriptive analyses were obtained, basing on the transcriptions and listening to the recordings, and informal observations were also made during the teaching of the songs.

Experiment Two

A diatonic song set to different texts was taught to three separate groups of children aged three to five in Hong Kong.
Effects of the pitch relationship in Cantonese songs

Group A: 91 Cantonese-speaking children (39 girls; 52 boys) from a local kindergarten who learnt the song with a Cantonese tone-mismatched text;

Group B: 63 Cantonese-speaking children (24 girls; 39 boys) from the same kindergarten who learnt the song with a Cantonese tone-matched text;

Group C: 56 English-speaking children (34 girls; 22 boys) from an international kindergarten who learnt the song with an English text.

After one month, I recorded the singing of each individual child on digital audio tapes with a DAT recorder. During the recording the children were allowed to sing freely, and neither a cue nor starting note was given.

With the use of a pitch to MIDI converter which was connected to a MIDI-interface, the digital information of the recordings was transformed into MIDI information. This was then entered into an Apple Macintosh computer for analysis and for sound generation purposes. Using the Digital Performer 1.4 program, frequency analysis was carried out.

By studying the pitch graphs and listening to the recordings, a note was identified for each word in the song for scoring. Criteria in the choice of notes were established in order to ensure the reliability of this process — either it was the longest note that the child sang for that word, or if it was a sliding note, the final pitch was selected. The singing of 194 children was analysed in total after eliminating those whose singing could not be processed by the computer (either it was not loud enough or was too breathy to be identified). The pitch data were represented in their MIDI frequency values.

Three aspects of singing were analysed quantitatively:

(a) Accuracy in intervallic relationships of the whole song
(b) Accuracy in melodic direction of the whole song
(c) Pitch accuracy of individual intervals

ANOVA was then performed on the results of (a) and Analysis of Deviance on (b) and (c) with Genstat 5 (Genstat 5 Committee, 1987) in order to test if the differences were statistically significant. Full account of all the possible interactions of three factors, group, age and sex, was taken. Afterwards, post hoc tests were also carried out in order to clarify the significant differences between the three groups.

To ensure the validity of the experiment, a number of further analyses was carried out. First, the total scores of error was re-calculated with a higher control of the experimental conditions in order to verify whether the results were robust. The control included eliminating the scores of all the tone-mismatched intervals in the tone-matched song as well as the tone-matched intervals in the tone-mismatched song; and then comparing the scores of the different phrases of the song which possessed different proportion of tone-matched and tone-mismatched intervals. Furthermore, by measuring the singing range of the children in singing the experimental song, which
serves as an indication of their singing ability, it was clarified whether the differences in children's singing ability had affected the results.

In addition, since the computer scores were the basis of most of the analyses, it was therefore important also to verify the external validity of these data. Correlation between the computer scores and the teacher ratings on part of the data was conducted in order to see whether the computer scores correspond to the teacher ratings which are the common assessment of children's singing in the real-life situation.

The qualitative issues raised in Experiment One were re-addressed with the assistance of computer-generated data and graphs using the computer software, Waves+/IESPS (Entropic Signal Processing Software) (Entropic Research Laboratory, 1993), which provides the details on intonation of children's singing. Moreover, evaluation of the singing of the texts of songs through listening to the recordings was performed and my informal observation of the learning processes were considered.

Findings

This study demonstrates that the effects of the pitch relationships between text and melody in young children's singing is a complicated issue, and the results did not emerge as predicted. The tone-matched and tone-mismatched songs were not performed significantly differently in terms of pitch accuracy. Furthermore, the diatonic song which was supposed to be most appropriately matched with English text turned out to be sung with lower accuracy in English than in Cantonese.

It appears that the differences in singing accuracy are not caused by text-melody relationship, but rather by children's educational background. The Cantonese-speaking children in Hong Kong have more training in singing tone-mismatched than tone-matched songs at school. Moreover, in comparison with the English-speaking children in this study, they are more familiar with the context of learning and recalling songs. However, it is also possible that the higher accuracy performed by the Cantonese-speaking children can be accounted for by their ability in tonal language. This was supported by some of the data demonstrating that the linguistic differences between Cantonese and English have led to the differences in children's singing style (the more detached pitches in the former and the more legato pitches in the latter) which were shown to be related to pitch accuracy. The results have also suggested that speaking Cantonese can help children to achieve higher accuracy in the singing of wider intervals like the major sixth and the perfect fifth than English-speaking children.

It is interesting to note that such effects of text-melody relationship on singing accuracy may vary across different sexes and ages in the singing of certain intervals. The boys in the Cantonese-speaking groups have been shown to be comparatively less affected by the text-melody relationship than the girls in the same group when singing the ascending major sixth. The three-year-old Cantonese children were more affected by this relationship in comparison with the four- and five-year-olds in the singing of the ascending perfect fourth.
Further qualitative considerations have shown that text-melody relationship is influential on other aspects of singing beside pitch accuracy. The linguistic tonal articulations were found in the singing of Cantonese texts, and the English children sang in a more legato manner than the Cantonese children. Furthermore, the English text was recalled best and the Cantonese tone-mismatched text was recalled the least efficiently. It was also suggested that Cantonese children are sensitive to the text-melody relationship as shown in their substitution of words, and they learn the tone-matched songs more efficiently than the tone-mismatched songs.

To conclude, the significance of text-melody relationship in songs should not be overlooked by music educators. The theoretical explanation to the present findings was that text and melody are integrated in children’s thinking. Past investigation of the integration hypothesis has been restricted to research on song recognition (Serafine, Crowder & Repp, 1984; Serafine, Davidson, Crowder & Repp, 1986; Crowder, Serafine & Repp, 1990; Morrongiello & Roes, 1990). This study extended this proposition to explain the children’s cognitive processes of song-learning and singing. It was argued that such integration exists in different aspects in the cognitive processes. Three aspects of text-melody integration were identified: children’s linguistic and musical abilities (the internal potential), the songs (the external stimuli) and children’s cognitive strategies (the internal response). Moreover, within each aspect, the degree of integration may vary under different circumstances.

One significant implication derived from this study is that music and language should be closely related in music education. In this way, not only the ability in one area will facilitate the ability in the other, but the learning of various qualities of sound can also be enhanced through both media as children can have a more diversified experience of those sound qualities in the context of both music and language. Furthermore, the higher pitch accuracy achieved by Cantonese-speaking children implies that pitch ability required in music can be acquired at early stages as children acquire their language. It seems that Cantonese children sing more accurately not only because they have more training at school, but more importantly, they have learnt about pitch discrimination in their language which can possibly be transferred to their singing skill. If the English-speaking children have the chance to be trained in their pitch ability early in life like them, they may sing as accurately as the Cantonese-speaking children. Therefore, music education in early childhood is very important.

REFERENCES


Genstat 5 Committee (1987) *Genstat 5* [computer program]. Harpenden, UK.


Background

With the increasing culturally diversified population in many countries of the world, cross-cultural research has gained a significant position in the current academic field. In the area of children’s singing research, the work of Buckton (1988), Moore and Kemp (1991), and Moore, Fyk, Frega and Brotons (1995/96) can be regarded as some of the most representative cross-cultural investigation. On the whole, these studies have illustrated that environmental influence is an important factor affecting children’s singing ability, and more specifically, the amount of enculturation and training in singing within their culture.

In fact, the issue of cultural differences in children’s singing is complicated and there are still many possible factors which have not been adequately explored. This paper focuses on language characteristics as a cultural factor affecting children’s singing. Blacking (1967) has documented how music and language are related in Venda children’s songs. In these songs, the metres are often determined by the spoken rhythm of the first word-pattern in their text and their melodies are closely related to the tones of ordinary speech. Similarly, Addo (1996) also demonstrated the influence of Fanti linguistic tones on the melodic contour of Ghanaian children’s play songs. Minami and Umezawa (1990) found that Japanese speech melody always appears in children’s improvisation, no matter whether they sing with words or with nonsense syllables. Fujita (1990) has observed ‘intermediate performance between talking and singing’ among the Japanese children. This style appeared not only in the singing of traditional songs, but also in talking rhythmically, both of which possessed a strong character of Japanese language.

I had conducted comparative research on Cantonese- and English-speaking children, as pitch has different functions in these two languages. In Cantonese, pitch serves a semantic function in the form of linguistic tones by determining the meaning of the words, which is unlike intonation in English. Such differences result in completely different pitch inflections of the two languages. It was found that language characteristics can possibly affect children’s perception of the expressive components in music and language (Chen, 1990) and singing accuracy (Chen-Hafteck, 1996) — Cantonese-speaking children had a different strategy in perceiving emotions in music and language from the English-speaking children, and they achieved higher pitch accuracy in singing than the English-speaking children. Therefore, in summary, the role played by language, particularly tonal languages, in children’s singing should not be undermined.
Assessment of children's singing

The reliability of measurements in singing has always been a concern for researchers. For example, many previous studies collected pitch data quantified to the nearest semitone, yet it has frequently been argued that such quantification may lead to some important minute details in the singing being ignored. This problem is even more pronounced in research involving young children's singing. This is often due to the pitches that young children sing during the early stage of their musical development not coinciding with the conventional scales used by adults (McKernon, 1979; Dowling, 1984).

The controversy between human-based and machine-based pitch analyses has been highlighted by Welch (1994). In fact, most of the past cross-cultural research (Blacking, 1967; Buckton, 1988; Minami & Umezawa, 1990; Fujita, 1990; Moore & Kemp, 1991; Moore, Fyk, Frega & Brotons, 1995/96) were based on human analyses. The human-based analyses offer us more information on the musical aspects of singing, based on human perception but they miss cut some details which may be significant to the research. On the other hand, the machine-based analyses give us precisely all the minute details of different properties of sound such as frequency, amplitude, and spectrum, yet they cannot infer musical meaning to the sound, a quality which is only possible through the human ear and mind. Walker's (1986; 1989) research is a good example demonstrating the advantage of using computer analysis in the study of cultural differences in singing.

I believe that there should be a balance between the two kinds of analyses. Human assessment is important to inform us the phenomena in singing which are significant to human ear. At the same time, scientific measurements can provide detailed and reliable data to support and verify the human judgement. The approach of the present study is to use the computer pitch data to verify and explain some human observations and statistical findings in the previous studies (Chen, 1992; Chen-Hafteck, 1994; 1996).

The present investigation

The aim of this study was firstly to clarify the relationship between language characteristics and children's singing styles, and secondly, to examine the effects of the pitch relationship between text and melody in Cantonese songs on children's singing. It was predicted that the differences between Cantonese and English languages and the pitch relationship between text and melody in Cantonese songs can exert some influence on children's singing styles.

Three children aged between four and five sang a diatonic song (a modified version of a Western children's song, 'Ten Green Bottles') with the same melody but set to three different texts. Children A and B are Cantonese-speaking, Child A sang the song with a Cantonese tone-mismatched text (where the linguistic tonal pattern mismatches the melodic contour) and Child B sang with a Cantonese tone-matched text (where the two match). Child C is English-speaking and she sang the song with an English text. The children's singing was recorded on digital audio tapes with a DAT recorder (Sony 419).
Languages and singing styles of children

During the recording the children were allowed to sing freely, and neither a cue nor starting note was given.

Analyses on amplitude and pitch was carried out using the computer software, Waves +/ ESPS (Entropic Signal Processing Software) (Entropic Research Laboratory, 1993). Two graphs were produced for each recording extract (see Appendix A1). The graph at the top shows the acoustic wave form and the graph below shows the fundamental frequency trace. The former is the acoustic correlate of loudness whereas the latter is the acoustic correlate of pitch. In order to compare the pitch singing of each corresponding note of the song, the time scale was modified. For instance, Child B sang much faster than Children A and C, and thus, the time scale of Child B’s graphs was expanded so they could be comparable with the graphs of the other two children.

Past studies have suggested that there were some differences in the singing styles of Cantonese- and English-speaking children (Chen-Hafteck, 1994); the overall articulation of linguistic tones in the song-text (in terms of pitch levels and inflections) could be shown only in tone-matched Cantonese songs, not tone-mismatched ones (Chen, 1992); and the detached singing style of Cantonese-speaking children helped them to achieve higher pitch accuracy over the English-speaking children (Chen-Hafteck, 1996). The present study attempted to verify these propositions with the aid of computer data.

Findings

When comparing the singing of Children A and B with Child C, some differences in singing styles related to language differences can be identified. By looking at the graphs of wave form or amplitude, we could observe how smooth or how discrete the sung pitches were. For instance, at Ie to Ih, we can see a large block of sound, indicating that there was no interruption in the singing of a few words of the text consecutively. This is because the text at that point did not have many consonants between words (‘flying in the’). In contrast, at Ila to Ile, there were several consonants in the text (‘But look up you can’) and so, we can see that the wave form was less crowded together.

Due to the individual tones for each Cantonese word as well as the presence of consonants in front of most of the words, Children A and B sang more discrete pitches than Child C. There were a few occasions where Child C tended to link some words together in her singing. Examples are at Ie to Ih, IIf to I1h, and IVi to IVk sung by Child C. On the contrary, in most of the singing of Children A and B, we can find that the words are quite distinctly separated. Therefore, all these occurrences may suggest that in general, English-speaking children tend to sing in a more legato manner whereas Cantonese-speaking children tend to sing in a more detached manner. This is likely to

---

1 Due to the page limit of the Conference Proceedings, it is not possible to include Appendix A which consists of four pages of graphs here. The Appendix will be distributed to the conference participants at the conference. Copies can also be obtained from the author.
Languages and singing styles of children

be related to differences between English and Cantonese. Thus, the observations reported in Chen-Hafteck (1994) has been supported.

Differences in the singing style of tone-mismatched and tone-matched songs by the two Cantonese children, Children A and B, have also been noted. Chen (1992) found that the articulation of linguistic tones appeared exclusively in tone-matched songs. But the present data has revealed that this issue is more complicated than that. When we look at the singing of words with rising tones, the overall articulation of the linguistic tones in the song-text (in terms of pitch levels and inflections) was clearly exhibited, but it appeared to be limited to the singing of the tone-matched song. It was evident that the rising tones were sung with an upward sliding pitch movement, as found at Id, Ilh, Ille and IVe in the singing of Child B. On the other hand, all the rising tones sung by Child A did not conform to this pattern. This is to say that the singing of the tone-mismatched song did not show the same articulation of rising tones.

However, the manifestation of clipped tones was different from that of rising tones. Clipped tones are short tones and therefore, they were sung as short notes, as illustrated in the singing of Child A at Id, Il and IIIe as well as in Child B at Ic and IIIi. But such a manifestation of clipped tones was not shown when the words fell onto long notes with a length of a crotchet or longer. Examples are found in the singing of Child B at Ib, IVe and IVc.

Therefore, it was evident that the articulation of rising tones is typically shown in the singing of tone-matched songs exclusively whereas the clipped tones are exhibited in both tone-matched and tone-mismatched songs. This is because the former is concerned with pitch and the latter concerns duration. Since tone-matched and tone-mismatched songs are characterized by their pitch inflections, whether a song is tone-matched or not can only have an effect on the manifestation of tones that are related to pitch. On the other hand, tone characteristics relating to duration can similarly affect both kinds of songs.

In Chen-Hafteck (1996), it was suggested that the higher pitch accuracy achieved by the Cantonese-speaking children may be due to language differences. A possible explanation is that the Cantonese language implies more discrete pitches when Cantonese words are sung in songs due to its tonal system, and discrete pitches may help children to sing wide leaps in pitch. This point has been illustrated in the present analysis.

Evidence showing that singing in a more detached manner can help in achieving pitch accuracy of wider intervals is found in the singing of ascending perfect fifth at IVd and IVe. Child A who had sung the two notes separately achieved a fairly accurate perfect fifth whereas Children B and C who had sung the two in a legato manner did not achieve this interval accurately.

In the singing of the ascending major sixth at IIIc and IIId where there was a large leap in pitch within a short rhythm (a quaver to a dotted crotchet), all the three children slid the pitch upwards. Both Children B and C did not reach the interval of a sixth. It was then suspected that the clipped tone at IIIb which made Child A cut short this note
before singing the wide interval of major sixth had helped her to achieve a higher accuracy than the other children. When we look at Children B and C, they did not have this break to give them a pause to prepare for the large interval. Thus, it again shows that singing in discrete pitches can help in pitch accuracy of large intervals.

It was also suggested that in singing the perfect fourth, the tone-matched condition could have helped in its accuracy (Chen-Hafteck, 1996). This has been demonstrated in the computer graphs. At IVa and IVb, the tones of the two words of text were the same, so Child A sang the two pitches quite closely, that is, less than a minor third, despite the fact that they were discrete pitches. Child B also sang in discrete pitches, and yet with the help of the linguistic tonal relationship, he had performed this interval accurately. As for Child C, she sang legato and did not perform this interval well.

**Conclusion**

The analysis of the acoustic aspects of singing offered detailed and objective information concerning the singing styles and the intonation patterns in singing. Such pitch data generated by computer has, in fact, facilitated us to understand and explain the phenomena in children’s singing which have been found in the past.

From the present investigation, it was suggested that there is a relationship between language characteristics and children’s singing styles, which can be one of the important factors explaining the cultural differences in children’s singing.

Moreover, the results imply that tone-matched Cantonese songs can encourage children to articulate the linguistic tones in their singing. This may suggest that tone-matched Cantonese songs are in a better position than tone-mismatched songs in developing children’s linguistic skills through singing. Furthermore, the finding that they help children to sing certain interval more accurately again illustrates that tone-matched Cantonese songs are perhaps good teaching materials for music teachers.

Another important finding in this study is that singing in a detached manner can facilitate children to sing more accurately large intervals such as perfect fifth and major sixth. Therefore, singing practice should include exercises to sing discrete pitches, especially for English-speaking children who do not sing in a detached manner naturally like Cantonese-speaking children. This method can possibly help to achieve higher accuracy in the singing of wide intervals.

**References**


Children's ability to identify two simultaneous melodies

Helga Rut Gudmundsdottir
Faculty of Music, McGill University

Introduction

The purpose of the present study was to investigate children's ability to hear two simultaneous melodies, the development of this ability with age, and whether alterations of the musical stimuli would affect the children's performance on this task.

When investigating perception of simultaneous melodies, the extensive literature on single-line melodic perception is relevant. It appears logical to presume that factors involved in perceiving a simple melody would be involved in the perception of two or more melodies. It has been shown that melodic perception improves with age (Hargreaves, 1986; Bartlett and Dowling, 1980; Costa-Giomi, Gudmundsdottir, and Demorest, 1996; Imberty, 1969; Trehub, Morongiello and Thorpe, 1985). Several factors have been found to affect melodic perception: tonality (Watkins, 1985; Dowling, 1991; Bartlett and Dowling, 1988), familiarity (Dowling and Fujitani, 1971; Dowling, 1978; Bartlett and Dowling, 1980; Hargreaves, Castelli and Crowther, cited in Hargreaves, 1986), length of the stimulus (Edworthy, 1985) and contour complexity (Cuddy, Cohen, and Mewhort, 1981). Recent research in melodic perception suggests that there is no single variable that controls the perception of melody (Dowling, 1994).

The added complexity of hearing two melodies instead of only one has not yet been determined. The findings of research in perceptual grouping and stream segregation is relevant in this regard because they indicate that an auditory stimulus is not perceived as merely the sum of its parts. Auditory stimuli form streams of perception, depending on variables such as frequency, pitch, timbre and loudness (see e.g. Bregman, 1990; Hartman & Johnson, 1991). For example, in a rapid sequence of tones that are of alternating high or low frequencies, the sounds are organized into two perceptual streams. This 'perceptual fusion' however, will not occur if the sequence is played at very slow speed (Bregman & Campbell, 1971). Bregman and Pinker (1978) further found that onset/offset synchrony is an important determinant of simultaneous effects and that tones that seem likely to have come from the same source (same timbre) are most often grouped into the same stream.

Although music listening and performance involves perceiving simultaneous events, little is known about children's perception of simultaneous melodies. Discriminative abilities of simultaneous events in music, such as melodies and timbres are not tested in most common standardized musical tests (see Boyle and Radocy, 1987; Gordon, 1965; Seashore et al., 1960). However, some tests measure the discrimination of simultaneous sounds (Wing, 1961; Bentley, 1966).

Zenatti (1969) found that children were more likely to recognize a theme in soprano and alto lines, and did so faster than in the bass line when the music was in four parts. Some studies showed that young children attended more to the upper melodic line than the lower one when presented with two-part music (Imberty, 1969; Zimmerman, 1971). Others did not find that young children necessarily attended to the upper melody in two-part music (Costa-Giomi and Pennycook, 1994).

Research has shown that children have difficulty discriminating simultaneous events in music (Serafine, 1981, 1988). In fact, young children have difficulty focusing on more than one aspect of a musical stimulus at once (Costa-Giomi, 1994a; Costa-Giomi, 1994b; Sims,
Older children can perform complex tasks, like focusing on relevant information while filtering out irrelevant information, but young children do not display the ability to do so. (Doyle, 1973; Maccoby & Konrad, 1966; Sergeant & Roche, 1973).

Serafine (1988) tested children's and adult's ability to recognize two newly learned tone sequences (4 beats). The subjects responded verbally to the question "would the melodies sound like this together." This proved to be a difficult task even for the adults (73% success) and the 5 - 8-year-olds performed at or below chance level. Gudmundsdottir (1995) conducted a similar study with 5-, 7-, and 9-year-olds using longer melodies (8 beats) which were both newly learned and familiar. The children scored significantly higher with the familiar melodies than the newly learned melodies. In fact, responses to the newly learned melodies were highly inconsistent. While the 9-year-olds were able to identify simultaneous familiar melodies quite accurately, the 5-year-olds tended to identify only the upper melody.

Children's performance in musical tasks varies significantly depending on the task's structure and design (Sergeant and Boyle, 1980). Children tend to receive higher scores on a given task when responding non-verbally than when responding verbally (Hair, 1977; Scott, 1978; Taebel, 1974; Webster and Schlentrich, 1982).

Method

Subjects

The children in this study were English speaking 1st (n =29), 3rd (n =22), and 5th-graders (n =22) (6-, 8-, and 10-year-olds). These children attended a public school in Montreal and received weekly music lessons by a music specialist.

Stimuli

Three melodies were used in the stimuli, two familiar and one unfamiliar. The familiar melodies were "The Barney Song" (B) and "Frere Jacques" (F). The unfamiliar melody (X) was composed for the purpose of this study. All three melodies were of the same length. Each test item consisted of two melodies played simultaneously in various combinations, as explained in Table 1. The three melodies were combined into six different melody pairs, and each pair was played in three timbre combinations: 1) piano / piano, 2) piano / trumpet, 3) trumpet / piano. Each pair was played once in every timbre-combination except for pairs (B+F) and (F+B) which were played twice in each timbre combination. Thus there were eight pairs in each combination and a total of twenty-four items in the test. The timbre combinations were presented in randomized order and so were the items within each combination. The length of each item was 16 seconds. There were 2 seconds of silence between items. The total duration of the test was 7.2 minutes.

Apparatus

The program that generated the stimuli and recorded the subjects' responses was developed with the Opcode's Max software system. The hardware consisted of a portable Macintosh computer, connected to a Roland Sound Canvas via MIDI and a Boss loud-speaker.

Procedure

Each subject was seated facing the computer with the loud-speaker at one side. Three keys of the computer keyboard were specially marked with stickers, symbolizing the three songs of the stimuli. After a short pre-test and training the subjects were instructed to listen to the
stimuli and to press the corresponding key as soon as they recognized a melody. Responses for each subject were automatically stored in a computer file. The computer recorded which keys were pressed for each of the items and the time it took to respond by counting the number of beats from the beginning of the melody until each response occurred. When the same key was pressed more than once during an item, only the first response was counted. The children could identify from 0 to 3 melodies but each test item consisted of only two melodies. A completely correct response to a test item should have two identifications, both of them correct.

Table 1: Timbre- and register combinations of the three melodies

<table>
<thead>
<tr>
<th>Item</th>
<th>Timbre 1</th>
<th>Timbre 2</th>
<th>Timbre 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(U): piano</td>
<td>(U): piano</td>
<td>(U): trumpet</td>
</tr>
<tr>
<td></td>
<td>(L): piano</td>
<td>(L): trumpet</td>
<td>(L): piano</td>
</tr>
<tr>
<td>1.1</td>
<td>B+F</td>
<td>B+F</td>
<td>B+F</td>
</tr>
<tr>
<td>1.2</td>
<td>F+B</td>
<td>F+B</td>
<td>F+B</td>
</tr>
<tr>
<td>1.3</td>
<td>B+X</td>
<td>B+X</td>
<td>B+X</td>
</tr>
<tr>
<td>1.4</td>
<td>X+B</td>
<td>X+B</td>
<td>X+B</td>
</tr>
<tr>
<td>1.5</td>
<td>F+X</td>
<td>F+X</td>
<td>F+X</td>
</tr>
<tr>
<td>1.6</td>
<td>X+F</td>
<td>X+F</td>
<td>X+F</td>
</tr>
<tr>
<td>1.7</td>
<td>B+F</td>
<td>B+F</td>
<td>B+F</td>
</tr>
<tr>
<td>1.8</td>
<td>F+B</td>
<td>F+B</td>
<td>F+B</td>
</tr>
</tbody>
</table>

Results

Three randomly selected children from each grade level took the test two times on two different occasions in order to establish the reliability of the measurement. A paired t-test performed on first scores vs. second scores indicated that the two were not significantly different (Mean = -1.44; SD = 3.21; T-value (-1.35); df = 8; p = .21).

For each test item the number of melodies identified were tallied, as was the proportion of correct responses. Most children identified two melodies while very few children identified zero or three melodies. Eighty three percent of all the items elicited two-melody responses while fifteen percent of them elicited one-melody responses. The frequency of two-melody responses was highest in grade 5 and lowest in grade 1. For two-melody responses, the proportion of correct responses was high in grade 5 (97%), lower in grade 3 (86%) and lowest in grade 1 (75%). The number of correct one-melody responses was lowest for 1st graders (95%), higher for 3rd graders and the 5th graders had all of them correct.

Statistical methods were used to examine the effects of age, gender, and contrasting timbres on the children's success in identifying two simultaneous melodies. The accuracy of responses and response speed were examined through analysis of variance (ANOVA). The effect of placement in register on the children's identification of melodies was examined using Chi-square analyses. An alpha level of .05 was used for all statistical tests.

Scores were significantly affected by Grade ($F(2, 58) = 13.098, p < .05$). Tukey post hoc analysis revealed that accuracy scores increased significantly from grade one to three and from grade three to five. Scores for items with two familiar melodies were higher than for the scores for items with one familiar and one unfamiliar melody ($F(7, 406) = 2.690, p < .01$). Timbre affected general scores significantly ($F(2, 116) = 2.90; p < .05$). Two melodies were correctly identified significantly more often when the trumpet played the lower melody and the piano the upper melody than when the melodies were played the other way around.

426
For further exploration of the data the correct two-melody responses and the correct one-melody responses were examined separately. Analysis of the correct two-melody responses showed that the older children (3rd and 5th-graders) responded significantly faster than the younger children (1st-graders) when speed was measured as the number of beats counted until both melodies were identified \( F(2, 772) = 12.803; p < .01 \). However, there was not a significant difference between grades on how long it took to identify one melody.

The children who correctly identified two melodies responded significantly faster to items that had two familiar melodies than to items with one familiar and one unfamiliar melodies \( F(5, 772) = 5.514; p < .01 \). The same children identified melodies in upper register significantly faster than melodies in the lower register \( (df = 825; T = 4.34; p < .01) \). Both timbre and register affected the order of identification. The trumpet melody was significantly more often identified first than was the piano melody \( (df = 2; \text{Chi-Square} = 120.7; p < .01) \). When the piano played in both registers the tendency was to identify the melody in upper register first (Figure 1).

Analysis of the correct one-melody responses showed that the identified melody was more than twice as often the upper melody than the lower melody. However, comparison between the three timbre combinations revealed that the lower melody was chosen significantly more often in combination 2 (u: piano / I: trumpet) where the trumpet played the lower melody, than in combinations 1 (u: piano / I: piano) and 3 (u: trumpet / I: piano).

Figure 1: Proportion of upper and lower melody identified first according to timbre

![Proportion of upper and lower melody identified first according to timbre](image)

**Timbre combinations**
- Upper first
- Lower first

- Timbre 1: u = piano / I = piano
- Timbre 2: u = piano / I = trumpet
- Timbre 3: u = trumpet / I = piano

**Discussion**

This study measured children's ability to attend to two melodies simultaneously. Nonverbal measurement allowed for collecting data on the speed and order of identification. As confirmed by statistical reliability measures, the computerized test procedure proved to be appropriate for the age groups involved.

A developmental age trend was found in the children's performance. The older children responded faster and more accurately than the younger ones. The younger children were more likely to identify only one of the two melodies than the older children. Interestingly, all the children identified one melody with a similar speed. The age difference emerged in the time they took to complete the identification of both melodies.

Of the two timbres used in the study, piano and trumpet, the children heard the trumpet timbre more readily than the piano timbre when both where played together. Perhaps children who could not focus easily on two melodies simultaneously attended to the melody with the more distinct timbre regardless of its register, although it is not clear why the trumpet would
be more distinctive than the piano. When the two melodies were played with the same timbre the children seemed to spontaneously attend to the upper melody.

The key to identifying two simultaneous melodies might lie in the ability to attend to the lower of the two melodies. The children who accurately identified both melodies took longer to identify the melody in the lower register than the melody in the upper register. Children who only identified one melody selected the one in upper register most often. However, timbre has possibly a stronger effect than register placement. This is suggested by the tendency of all the children to favor a trumpet melody regardless of register.

Further research is needed to understand the ability to hear simultaneous events in music and how this ability develops in children. Future research should explore the effect of different timbre characteristics on perception of stimuli with more than one timbre.

Acknowledgments

This project was supported by a grant from McGill University awarded to Dr. Eugenia Costa-Giomi.

References


Children's judgments of structural patterns and emotional expression in music

Dagmar Bergman and Alf Gabrielsson, Department of Psychology, Uppsala University, Sweden

Children's ability to discriminate among pitches, rhythm patterns, melodic contours and other structural variables has been the subject of many studies; for reviews see Gembris (1995), Hargreaves (1986), or Sloboda (1985). Their ability to recognize emotional expression in music is still little investigated, but recently a number of studies have appeared dealing with this issue. Kratus (1993) studied 658 6–12 year olds' ability to recognize the dimensions happy – sad and agitated – calm. He used short excerpts from J. S. Bach's Goldberg Variations, performed on the piano by Glenn Gould, and the children should answer by encircling the picture of a face corresponding to the emotion supposed to be expressed. Children of all age groups and both genders were highly consistent in their interpretation. There were no significant age or gender differences, but children were more consistent concerning the distinction between happy and sad than regarding the distinction between excited and calm.

Dolgin and Adelson (1990) used 16 composed melodies, judged by adults as being either happy, sad, angry or frightened–sounding and performed on the viola or sung by a soprano voice, to study the recognition of these expressions by 4–, 7– and 9–year–olds and by college students. They responded by choosing among drawings of faces (4–year–olds) or both drawings and names of the emotions (7– and 9–year olds). Happy melodies were best recognised, followed by sad, angry and frightened melodies in that order. On the whole, recognition increased with increasing age but in different ways for different emotions and for different performances (viola, song).

Terwogt and van Grinsven (1991) had 5– and 10–year–old children and adults to judge the emotional expression – happiness, sadness, anger, fear – in short classical musical extracts by choosing one of four schematically presented facial expressions. Recognition increased with increasing age, and happiness and sadness were recognised better than anger and fear. The latter were often confused with one another, especially by the 5–year–olds, and possible reasons for that were discussed, for instance, that the emotion of aggression produces fear. In a separate study 30 adults rated 100 words (in Dutch) with regard to how well the emotion expressed by each word might be expressed in music. Words denoting happiness, sadness and anger usually got high ratings, whereas mood–indicators associated with fear got lower ratings.

Giomo (1993) studied children, age 5 and 9, regarding their sensitivity to mood in music. Twelve short pieces from the classical repertoire were judged by the children using drawings of faces presumed to illustrate various degrees of three dimensions: intensity – softness, pleasantness – unpleasantness, and solemnity – triviality. There was no general significant difference among the two age groups. Girls performed better than boys at both age levels. Among the nine–year–olds children from homes with high socio–economic status performed better than children from homes with middle or low socio–economic status. Unexpectedly, children with no home musical experience performed better than children with such experience.

Trehub (1993) found that children 4 to 8 years of age judged melodic lines with high pitch and fast rate to be happy and melodic lines with low pitch and slow rate to be sad. Ascending
sequences led to happy judgments, descending to sad judgments. The answers were given by rotating a pointer to a sad, a happy, or a neutral looking face.

These studies indicate that young children are able to recognise emotional expressions, such as happiness, sadness, anger and fear in music, the former two easier recognisable than the latter two. However, there are some contradictory results concerning the effects of age and gender. This may be due to differences in the selection of music, methods for judgment and others.

Children's musical ability has thus primarily been studied with regard to structural features (pitches, intervals, rhythms, melodic contours etc) and only recently and to less extent concerning recognition of emotional expression. However, there seem to be no studies in which these two categories of abilities are tested within the same investigation. The purpose of the present study was to investigate the relation between children's ability to judge tonal and rhythmic structures as well as their ability to recognise emotional expressions in music. Possible gender differences should be studied as well.

METHOD

Twenty-one 9-year-old children, 10 boys and 11 girls, were tested using Gordon’s (1979) Primary Measures of Music Audition (PMMA). The results were compared with the results on an emotion test. A control group of ten music teachers also made the tests.

Material

The PMMA (Gordon, 1979) was used to examine the ability to compare the structures of tonal and rhythm patterns. It consists of two subtests, Tonal and Rhythm. Each part consists of 40 items plus 4 practice examples for the Tonal test and 2 practice examples for the Rhythm test. In both tests the child listens to a pair of tape-recorded short sequences, tonal or rhythmic, realized with a synthesiser, and has to decide if they sound the same or different. If they sound the same the child will encircle, on the answer sheet, a picture of two faces that look the same. If the phrases sound different the child encircles the picture of two different faces.

The Emotion test consisted of six short melodies of four or eight bars which were composed for this investigation. The melodies were played in four different ways on the violin by the composer to express the emotions happy, sad, angry and scared. All 24 performances (six melodies x four expressions) were recorded on tape. The answer sheets had the four emotions represented by four schematically drawn faces for each of the 24 items.

Procedure

The PMMA was mainly administered in accordance with the instructions in the PMMA manual (Gordon, 1979) providing adequate instructions and explanations of the task. The 21 children were divided into three groups of seven children in each group. The two tests were administered at separate occasions separated by one week, first the Tonal test and then the Rhythm test. The atmosphere was calm, peaceful and enjoyable. The children seemed to enjoy the situation.

The Emotion Test was conducted in the afternoon of the same day as the Rhythm test. It was first checked that the children could recognise the intended facial expressions by asking them to verbally express which of the faces illustrated the specific emotions: happy, sad, angry and scared. Then the children listened to one version of a melody, which was supposed to express one of the four emotions. The tester asked: "Do you think the lady who is playing this song is happy, sad, angry or scared?" The children were asked to encircle a face expressing the
emotion that they thought would be appropriate. Not until it was certain that all the pupils had understood how to cope with the task the whole emotion test of 24 items was presented. It took about 40 minutes to finish the test.

The control group of 10 music teachers made all three tests in a single session.

RESULTS

Maximum possible score was 40 in the Tonal and Rhythm test, 24 in the Emotion test. All children but two scored 30 or higher on the Tonal test and all but one on the Rhythm test. Fourteen children scored 18 or higher on the Emotion test. One girl reached the maximum score (40) on the Tonal test and also had the highest attained score (37) on the Rhythm test but got among the lower scores (18) on the Emotion test. The situation was quite opposite for another girl, who scored much lower than all other children on the two PMMA tests (24 and 21) but had the highest attained score (22) of all children on the Emotion test.

The mean values for boys and girls were almost the same both for the Tonal test (M = 34.2 for boys, 34.0 for girls) and the Rhythm test (M = 33.4 for boys, 33.1 for girls). In the Emotion test boys’ mean value (19.5) exceeded that for girls (18.0). This difference almost reached the conventional .05 significance level, \( t = 1.92, \text{df} = 19, p = .069 \). The means for the music teachers were very close to the maximum score in all three tests.

According to Gordon (1979) the correlation between the two PMMA tests should normally be approximately .50, which points to a shared variance of 25%. In this study the correlation between the two tests for the whole group was \( r = .61 \), which is roughly in agreement with Gordon's statement. However, the correlation was very different for boys and girls separately. For girls it was as high as .87, whereas for boys it was in fact slightly negative, -.18.

Results on the Emotion test

Maximum result on the Emotion test was 22 the result of a boy with fairly good results on Tonal as well as on Rhythm test but also of a girl with the lowest scores on both PMMA tests. Another surprising result is perhaps that a girl with top scores on both PMMA tests reached only 18 points on the Emotion test. Lowest scores on the Emotion test occurred for two girls with very good scores on the Tonal and Rhythm tests.

Table 1 shows the percentage correct responses for boys and girls across all melodies. It is obvious that Scared was much harder to recognize than the other emotions. However, the percentage obtained for Scared (57%) is still much higher than what would be expected to happen by chance alone, that is, 25 percent (pure guessing among four alternatives). Of course, the results varied among the different melodies; this is not discussed here.

Table 1. Percentage correct responses for boys and girls across all melodies

<table>
<thead>
<tr>
<th>Group</th>
<th>Happy</th>
<th>Sad</th>
<th>Angry</th>
<th>Scared</th>
<th>Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys</td>
<td>88.33</td>
<td>91.67</td>
<td>86.67</td>
<td>56.67</td>
<td>80.84</td>
</tr>
<tr>
<td>Girls</td>
<td>78.79</td>
<td>84.85</td>
<td>80.30</td>
<td>57.58</td>
<td>75.38</td>
</tr>
<tr>
<td>Means</td>
<td>85.56</td>
<td>88.26</td>
<td>83.49</td>
<td>57.13</td>
<td>78.11</td>
</tr>
</tbody>
</table>
The distribution of errors in recognition is shown in Table 2. The intended emotion appears in the rows and the perceived emotions in the columns. Both happy and sad can obviously be misinterpreted as scared. Angry is often wrongly recognised as happy. These three emotions have each one main alternative to the correct identification. This is not the case, however, when it concerns the emotion scared, for which the errors spread over all the other emotions with a higher number for happy.

**Table 2. Percentage Error Distribution.**

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Happy</th>
<th>Sad</th>
<th>Angry</th>
<th>Scared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
<td>-</td>
<td>2.38</td>
<td>1.59</td>
<td>10.32</td>
</tr>
<tr>
<td>Sad</td>
<td>1.59</td>
<td>-</td>
<td>0.00</td>
<td>10.32</td>
</tr>
<tr>
<td>Angry</td>
<td>15.08</td>
<td>0.00</td>
<td>-</td>
<td>0.79</td>
</tr>
<tr>
<td>Scared</td>
<td>20.63</td>
<td>11.90</td>
<td>10.32</td>
<td>-</td>
</tr>
</tbody>
</table>

**Correlations among the tests**

For the whole group the Tonal and Rhythm tests were positively inter-correlated \((r = .61)\). However, the Emotion test was negatively correlated both with the Tonal test \((- .24)\) and especially with the Rhythm test \((- .56)\). These negative correlations were more pronounced for the girls \((- .49\) and \(- .69\), respectively) than for boys \((.04\) and \(- .65\), respectively).

A factor analysis (principal components, varimax rotation) on the correlations for the whole group resulted in two factors accounting for 92.6 percent of the total variance. Factor I had the highest loadings in the PMMA tests but close to zero loading for the Emotion test. Factor II was bipolar, contrasting the Rhythm test against the Emotion test.

**DISCUSSION**

The results show some similarities to earlier investigations on children's recognition of emotional expression, especially that the emotion scared is more difficult to identify than others. With regard to gender differences the result in this study suggested that boys were slightly better than girls concerning recognition of emotional expression in the music examples used.

The correlation between 9-year-old children's ability to compare tonal and rhythmic patterns and to recognize emotional expressions in music was negative, especially regarding the correlation between the Rhythm test and the Emotion test. Of course, this result may depend on the specific tests used in this study and replications using other measures are necessary in order to see if this result will still be valid. We found some items in the PMMA Rhythm test doubtful as test of rhythm discrimination. For instance, no one of the 21 children and the 10 music teachers were able to recognise a difference between the two parts of items no 21 and 29, for which the correct answer should be "different" according to the test manual. In notation the the two parts in the respective items look different (see Figure 1), but the auditory impression is that they are the same. This impression is strengthened by an inspection of the amplitude envelopes of the respective patterns, realized by a sampling system; no evident difference between the parts could be seen in this analysis. Item no 28 is another dubious point, see Figure 1. The sounding result is a single tone of the same duration in both parts of the item. Of course, the correct answer is "same", but comparing two single tones can hardly be considered as a test of rhythm discrimination.
Figure 1. Items nos 21, 29 and 28 of the Rhythm test as notated in the PMMA manual.

Acknowledgment
This research was supported by the Bank of Sweden Tercentenary Foundation.

References


The Effects of Expressive Variation in Dynamics on the Musical Preferences of Elementary School Students

Vernon Burnsed
Virginia Polytechnic Institute and State University

Although expressive performance is widely endorsed, research into the effects of expressive performance in music education is virtually nonexistent. If it can be demonstrated that expressive performance has a positive effect on the musical preferences of students, the implications for music educators will be significant.

Expressive performance was first studied by Seashore (1938) who believed that "The medium of musical art lies primarily in artistic deviation from the fixed and regular" (p. 29). Seashore found that performers utilized much variation in intensity and duration to achieve expressive effect. More recent research has substantiated Seashore's findings (Friberg, Sundberg, & Fryden, 1987; Shaffer & Todd, 1987; Gabrielsson, 1987; Sundberg, Askenfelt, & Fryden, 1983). These studies, which have been primarily analytical, reveal that performers manipulate dynamics and duration to achieve expressive effect. As a general rule, dynamics increase and decrease with the rise and fall of the melodic contour. Duration varies in an opposite manner; note values decrease as the contour rises and increase as the contour falls. Duration increases also delineate phrases.

LeBlanc (1987) has proposed a theory of musical preference decision "... based upon the interaction of input information and the characteristics of the listener, with input information consisting of the musical stimulus and the listener's cultural environment" (p. 139). LeBlanc identifies physical properties, complexity, referential meaning, and performance quality as components of the musical stimulus that interact and influence the musical preference decisions of listeners. Dynamics have been shown to be related to mood, expressive effect, and performance evaluation, (Ortmann, 1927; Grundlach, 1935; Seashore, 1938; Gordon, 1965; Burnsed, Hinkle & King, 1982). Given LeBlanc's theory (1987), it seemed logical that dynamics may also influence preference decisions.

The purpose of this study was to determine the effects of expressive variation in dynamics on the folk song preferences of elementary school students. It investigated the relationship between two of LeBlanc's (1987) input variables, the physical properties and the performance quality of music, and musical preference decisions. Burnsed and Sochinski (1995) found that a significant proportion of middle school band and choir students preferred versions of folk songs with expressive variation in dynamics to those in which the dynamics were held constant. This was not unexpected since those students were somewhat select—they attended a music camp—and they had received performance group training where dynamic variation may have been taught as a part of performance technique. The present study sought to determine if expressive variation in dynamics might affect the preferences of the typical elementary student who attends music class only once or twice a week.
Method

Two studies were conducted. In the first a Macintosh LC computer with an accompanying Yamaha WT 11 Wind Tone Generator and head phones was placed in the general music classrooms of two suburban elementary schools in southwest Virginia. Over a period of approximately six months 123 grade K-5 students completed Music Impression Expression (Bishop, 1993), a preference test designed to measure the effect of expressive variation in dynamics on the folk song preferences of participants. The test was administered individually by the computer.

Music Impression Expression is a Hypercard (1991) stack written in the Hypertalk programming language. Two versions each of ten folk songs are included in the stack. For the non-expressive versions, all attacks, releases, decays, and volumes are set equally. For the expressive versions, the dynamics vary to follow the rise and fall of the musical phrase. A flute tone without vibrato was selected as the timbre for the folk songs.

The songs were selected from 150 American folk songs to sing read and play (Erdei, 1974). Melodies were selected for moderate complexity and varied contours. The songs were the following: The Bell Cow; Chase the Squirrel; Chickalileelo; Who Killed Cockrobin; Oh Fly Around My Pretty Little Miss; Jim Along Josie; Old Lady Sittin in the Dining Room; Pourquoi; Sailing Over the Ocean; and Sourwood Mountain.

The results of the first study were inconclusive. Reliability coefficients were low and the students did not appear to choose expressive or non-expressive versions of the folk songs in a meaningful pattern. There were no consistent effects for expressive dynamics or for the age of the students. It was speculated that perhaps the computer administration of the test was a confounding factor. The elementary school students in the first study worked semi-independently and they may not have been able to follow computer given instructions or complete tasks as purposefully as they would have in a teacher directed group administration. It was decided that a group administration of the preference test would be utilized in the second study.

In the second study participants were students from the first through the fifth grades at two urban elementary schools in southwest Virginia. The communities represented by the two schools were both suburban and inner city with a variety of ethnic groups of various income levels. The students were taught by the same music teacher for 30 minutes a week in general music.

Music Impression Expression (Bishop, 1993) was again utilized to measure the effect of expressive variation in dynamics on the folk song preferences of the participants. This time the teacher administered the test with a Macintosh LC computer, a Yamaha WT 11 Wind Tone Generator, and two Yamaha MS 101 powered monitor speakers. Each participant was given an answer sheet and the music teacher read the instructions which were similar to those recommended by Sims (1987). The teacher guided the participants through the answer sheet as each example was performed by the computer. After
approximately 3 weeks the students at both schools had completed the preference test. At School A 172 participants completed the test; at School B there were 143 participants.

Results

Reliability coefficients (Cronbach's Alpha) for Music Impression Expression were computed for the total participant population at each school. They were the following: School A = .53 ; School B = .51. Given the age of the participants and the nature of the test, these coefficients were deemed acceptable. Z tests to determine if a significant proportion of students by grade level preferred one version of each folk song over the other did not reveal a consistent age effect. Although the higher grade levels appeared to achieve more significant Z scores at School B, this was not the situation at School A where significant Z scores were more evenly spread among the grades.

Since there appeared to be no meaningful effect of age for both populations, the data were collapsed at each school and Z tests conducted on each song for the total participant population at each school. Table I reports the proportion of expressive choices and the Z scores for each song at the respective schools. Both schools significantly preferred the non-expressive version of song 1 and were undecided about songs 6 and 8. For all the other songs, at least one of the schools significantly preferred the expressive version.

Discussion

The conclusions drawn from this study are somewhat limited by the population. Students were from intact music classes in two elementary schools in southwest Virginia. Given this limitation, the results of this study indicate that expressive dynamics have a positive effect on the folk song preferences of elementary school students. Although the results of the first study were inconclusive, in the second study a significant proportion of first through fifth grade students chose the expressive versions of 7 out of 10 folk songs. These results support the findings of Burns and Sochinski (1995) and underscore the importance of expressive performance in music education.

It appears that a contributing factor in the results of the first study was the computer administration of Music Impression Expression. Although the computer administration of the preference test had proven successful with older students (Burns and Sochinski, 1995), this was not true for elementary students. Elementary students working semi-independently may not follow computer given instructions or complete tasks as purposefully as students given directions by a teacher.

An important finding of this study was the non-effect of age or grade level on the students' preferences for the use of expressive dynamics in folk songs. Although there were differences between grade levels, these differences were not consistent nor indicative of a systematic effect for age. The majority of the students in this study were able to listen to and compare two versions of the same folk song and make a choice based upon
the expressive manipulation of dynamics. If the results of the current study are supported by future research, it appears likely that elementary students perceive expressive nuance in music and respond to it in a positive manner. This should have significant implications for the practice of elementary music education.

Table 1

**Proportion of Expressive Choices & Z Scores by School for each Folk Song**

<table>
<thead>
<tr>
<th>Song</th>
<th>School A</th>
<th></th>
<th>School B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prop.</td>
<td>Z</td>
<td>Prop.</td>
<td>Z</td>
</tr>
<tr>
<td>1</td>
<td>.36</td>
<td>-3.81*</td>
<td>.37</td>
<td>-3.20*</td>
</tr>
<tr>
<td>2</td>
<td>.62</td>
<td>3.30*</td>
<td>.60</td>
<td>2.65*</td>
</tr>
<tr>
<td>3</td>
<td>.76</td>
<td>8.30*</td>
<td>.67</td>
<td>4.56*</td>
</tr>
<tr>
<td>4</td>
<td>.54</td>
<td>1.22</td>
<td>.63</td>
<td>3.38*</td>
</tr>
<tr>
<td>5</td>
<td>.72</td>
<td>6.67*</td>
<td>.70</td>
<td>5.41*</td>
</tr>
<tr>
<td>6</td>
<td>.50</td>
<td>0</td>
<td>.54</td>
<td>1.09</td>
</tr>
<tr>
<td>7</td>
<td>.69</td>
<td>5.64*</td>
<td>.53</td>
<td>.75</td>
</tr>
<tr>
<td>8</td>
<td>.52</td>
<td>.61</td>
<td>.46</td>
<td>-.75</td>
</tr>
<tr>
<td>9</td>
<td>.59</td>
<td>2.48*</td>
<td>.66</td>
<td>4.16*</td>
</tr>
<tr>
<td>10</td>
<td>.72</td>
<td>6.45*</td>
<td>.60</td>
<td>2.47*</td>
</tr>
</tbody>
</table>

* p <.05
References


Sensitivity to pitch change direction in first grade children

Janina Fyk
Fryderyk Chopin Academy of Music, Okólnik 2, 00-368 Warszawa, Poland

Jerzy Krystecki
State Higher School of Humanities and Education, 3 Maja 7, 99-400 Łowicz, Poland

Abstract

The purpose of the research was to find out if there is an interrelationship between the direction of tone pitch change and the ability of seven-year-old children to perceive pitch change. One hundred and ninety seven (197) first grade pupils from four country primary schools took part in the research. A Pitch Discrimination Aptitude Test (Fyk, 1991) consisting of 5 series each containing 12 tasks preceded by three practice tasks, was used in the research. Each task consisted of two tones: a standard tone and another tone which was to be compared with the standard. In half of the tasks there was a pitch change in the second tone, while in the remaining six tasks the pitch of both tones was the same. The number of upward and downward pitch changes was identical - three of each in each series. Within a given set the size of pitch change in the second tone of the pairs was constant and amounted to: 75 cents in series I, 60 cents in series II, 45 in series III, 30 in series IV, and 15 cents in series V. The test material was composed of tones of an oboe-like timbre produced on a YAMAHA synthesizer and recorded.

The hypothesis that the direction of the tone pitch change has an influence on children’s ability to perceive pitch change was tested. The children recorded their responses on test answer sheets. In order to define the influence of the direction of the pitch change on children’s ability to perceive change, the strength of the dependence (r Pearson) between correct responses in the case of upward pitch change and downward pitch change was calculated. Test t Student was carried out to establish whether the difference obtained was statistically significant. The results obtained in this work only partly confirm the above hypothesis. In fact, the direction of the tone pitch change only influenced the perception of changes amounting to 75 and 30 cents. This phenomenon is explained on the basis of categorical perception. When the pitch change amounts to 30 cents upward, a break-up in the category of prime occurs more often than in case of the same downward pitch change. It seems that the interval produced through changing the tone pitch by 75 cents downward assures to a greater extent the character of minor second than the interval produced through identical pitch change upward.

INTRODUCTION

Music psychologists put forward a number of concepts of musicality structure which show that tone pitch hearing is one of the fundamental components of musicality. Pitch hearing consists in the ability to identify the difference in pitch between two tones. The ability to detect a change in the pitch, i.e. sensitivity to pitch differences, is defined by the pitch difference threshold which is the perceived least difference in pitch between two tones.
Literature on the research on the development of child’s hearing sensitivity shows that this development takes place along with growing up.

Pitch hearing tests for early-school age children were carried out by a number of researchers. According to Gilbert (1893), six-year-olds can distinguish 3/8-tone intervals, seven-year-olds - 1/3, eight-year-olds - 1/4, and nine-year-olds - 1/8, i.e. 25 cents. According to Vetlugina (1963), 83% of pre-school age children can distinguish pitch of tones. On the basis of the research conducted by himself and by other scientists, Teplov (1952) stated that the average pitch difference threshold varies from 6 to 40 cents. Lower threshold is found with musically trained subjects, e.g. for music school pupils playing violin it reaches 6 to 11 cents, for pianists - 12-21 (Blagonadierrezyna, 1940). However, these results refer to persons of over ten years of age. Yet Bentley (1966), testing seven-year-olds, proved that some of them could distinguish 1/4-tone difference in pitch, while research made by Duell and Anderson (1967) shows that 59% of six-year-olds can distinguish 1/2-tone difference and 20% of eight-year-olds - 1/6. Recent research on the pitch difference threshold, using the assumptions of the theory of signal detection, proves that this threshold with six- and seven-year-olds can vary from 20 to 30 cents (Fyk, 1993; Fyk, Krystecki, 1994).

The ability to distinguish tone pitch by preschool children follows the capability for understanding the notion of „pitch” (Buckton, 1983; Fyk, 1985). In literature on the research of predominantly the ability to discriminate pitch change direction, we can find Williams’ tests (1977) in which a group of 32 seven-year-olds took part and the research conducted by Hair (1977). In both cases, however, the tone pitch change led immediately to the change of musical interval; in fact - through the addition of a third and a fourth tone - it involved recognizing the direction of melody’s flow.

The results obtained by the authors mentioned above, despite a number of similarities, show a certain degree of divergence, being the consequence of adopting different goals of research, methods, and instruments. They cannot be the basis for specifying whether the direction of pitch change affects the perception of pitch change. Therefore, the goal of this paper is to specify the effect of the direction of pitch change on child’s ability to detect the pitch change.

2. SUBJECTS

The research involved 197 first-grade pupils from four country primary schools of Mazowsze region. The tests were conducted in classrooms. The group comprised 115 girls and 92 boys. The average boys’ age was 7 years and 4 months, and the girls’ - 7 years and 3 months. Before starting school all the children participated in kindergarten class for six-year-olds, the so-called „0” group. None of them ever attended music school, played any instrument, or took part in any form of musically developing classes.

3. METHOD

The research made use of Pitch Discrimination Aptitude Test (Fyk, 1991) that was devised and made as a research test at The Laboratory of Musical Acoustics of the Fryderyk Chopin Academy of Music in Warsaw.

It consisted of five twelve-task series following three trial tasks. Each task consisted of a pair of tones, the first one of each was the model tone, the second one - a comparative tone, higher or lower from the model one. Six tasks of each series involved a change of the second tone,
while in the remaining six tasks both tones were of the same pitch. The number of tasks including up-pitch change was the same as for the down-pitch change - three tasks in each series. Within each series the degree of the second tone pitch change was constant and amounted to 75 cents in series I, 60 cents in series II, 45 cents in series III, 30 cents in series IV, and 15 cents in series V. The tones were of oboe-like timbre generated with a YAMAHA synthesiser, recorded on a DAT tape recorder.

Before commencing the tests all the children took part in two meetings, during which they were instructed how to fill in special test forms and the following notions: „pair” of tones, „the same” tones, „higher” second tone, „lower” second tone were explained to them. The children were asked to tick either pairs of identical pictures after they heard tones of the same pitch, or different pictures if they heard tones of different pitch.

The hypothesis tested was whether *pitch change direction affects pitch discrimination ability by the tested group of children.*

4. RESULTS

Altogether 11820 responses were obtained - 2364 for each series. Table 1 presents the percentage of correct responses in case of detecting the pitch change of the second tone in the *Pitch Discrimination Aptitude Test.*

<table>
<thead>
<tr>
<th>Magnitude of pitch change</th>
<th>Boys (N=82)</th>
<th>Girls (N=115)</th>
<th>Total (N=197)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
</tr>
<tr>
<td></td>
<td>↓</td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>I 75</td>
<td>76</td>
<td>70</td>
<td>73</td>
</tr>
<tr>
<td>II 60</td>
<td>83</td>
<td>71</td>
<td>76</td>
</tr>
<tr>
<td>III 45</td>
<td>66</td>
<td>51</td>
<td>58</td>
</tr>
<tr>
<td>IV 30</td>
<td>68</td>
<td>65</td>
<td>66</td>
</tr>
<tr>
<td>V 15</td>
<td>55</td>
<td>53</td>
<td>54</td>
</tr>
<tr>
<td>X</td>
<td>70</td>
<td>62</td>
<td>65</td>
</tr>
</tbody>
</table>

Table 1. The percentage of correct answers in the Test with respect to the upward (↑) and downward (↓) pitch changes.

As it results from the data obtained by the whole group of subjects and shown in the last two columns of Table 1, in a particular series a similar number of correct responses was obtained, both for up-pitch and down-pitch change. Only in case of 30-cent pitch change the up-pitch change was better discriminated (66% correct responses) than the down-pitch change (50% correct responses). Taking into account the sex of children it appears that boys are better at discriminating 75-cent down-pitch change of the second tone (82%) from the same magnitude of up-pitch change (76%). In discriminating 30-cent up-pitch change the boys got 68% correct responses and 57% in case of down-pitch change. In comparison with remaining magnitudes of pitch change, the number of correct responses for both directions of change was similar. Just like the boys, the girls are also better at discriminating 75-cent down-pitch change of the second tone than its up-pitch counterpart. They also more frequently discriminate 30-cent up-pitch change (65%) from the same down-pitch change (44%).
In order to define the effect of the second tone pitch change direction on the pitch change discrimination aptitude, the value of correlation coefficient between correctly discriminated upward and downward changes was calculated. Before calculating Pearson's $r$, the assumptions, necessary for employing linear measure of the relationship, were checked. The results of the calculations are shown in Table 2.

<table>
<thead>
<tr>
<th>Series</th>
<th>Boys (N=82)</th>
<th>Girls (N=115)</th>
<th>Total (N=197)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r$</td>
<td>$r$</td>
<td>$r$</td>
</tr>
<tr>
<td>I</td>
<td>.72</td>
<td>.67</td>
<td>.69</td>
</tr>
<tr>
<td>II</td>
<td>.53</td>
<td>.55</td>
<td>.55</td>
</tr>
<tr>
<td>III</td>
<td>.37</td>
<td>.41</td>
<td>.42</td>
</tr>
<tr>
<td>IV</td>
<td>.29</td>
<td>.43</td>
<td>.37</td>
</tr>
<tr>
<td>V</td>
<td>.34</td>
<td>.42</td>
<td>.39</td>
</tr>
</tbody>
</table>

Table 2. The values of Pearson's correlation coefficient ($r$) for the detection of upward and downward pitch changes in the Test. Level of statistical significance $\alpha = .01$.

The correlation coefficient reaches the highest value for 75-cent pitch change and it points, in case of both boys and girls, to a moderate correlation between detecting pitch change and its direction. A moderate, although lower, correlation between the ability to discriminate pitch change and its direction is also observed, again both for boys and girls, at 60-cent pitch change. In case of perceiving 45-, 30-, and 15-cent upward and downward pitch changes the $r$ coefficient values are roughly similar and point to a low correlation, no matter what sex. In order to find out whether the differences in the ability to discriminate upward and downward pitch changes of identical magnitude are statistically significant, a Student $t$ test was carried out. The results are shown in Table 3.

<table>
<thead>
<tr>
<th>Series</th>
<th>Boys (N=82)</th>
<th>Girls (N=115)</th>
<th>Total (N=197)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t$</td>
<td>$p$</td>
<td>$t$</td>
</tr>
<tr>
<td>I</td>
<td>-2.35</td>
<td>.02</td>
<td>-2.37</td>
</tr>
<tr>
<td>II</td>
<td>.64</td>
<td>no-s</td>
<td>-.48</td>
</tr>
<tr>
<td>III</td>
<td>-.09</td>
<td>no-s</td>
<td>1.18</td>
</tr>
<tr>
<td>IV</td>
<td>2.45</td>
<td>.02</td>
<td>6.27</td>
</tr>
<tr>
<td>V</td>
<td>.00</td>
<td>no-s</td>
<td>.57</td>
</tr>
</tbody>
</table>

Table 3. The values of Student's statistics ($t$) for mean differences and independent samples under detection of given pitch changes in the Test. $p$ - value of probability of $t$ statistics.

The data in Table 3 indicate that for the tested group of children, independently of sex, those differences are statistically significant in case of 75- and 30-cent pitch changes and insignificant for 60-, 45-, and 15-cent pitch changes.

5. CONCLUSIONS

The results obtained in the present research only partially support the adopted hypothesis about the effect of pitch change direction on the ability to perceive tone pitch change.
Specifically, in reference to the tested group of seven-year-olds, this hypothesis was verified in case of the biggest 75-cent pitch change, and the second smallest 30-cent pitch change, whereas the pitch change direction had no influence on the ability to detect the smallest 15-cent pitch change, and the changes of 60 and 45 cents. This variation in pitch change discrimination aptitude dependent on the pitch change direction, or the absence of such variation was observed both in the group of boys and the group of girls, which suggests similar perception tendencies in both sex groups.

On the basis of the percentage of correct responses, showing the ability to detect five magnitudes of pitch change in all the test series, a conclusion may be arrived at that both boys and girls are better at perceiving upward pitch changes than the downward ones. It is, however, significant that in case of 75- and 30-cent pitch changes (those that confirmed the hypothesis concerning the influence of pitch change direction on the ability to discriminate pitch change) contrary perception tendencies were observed. Thus, in case of 75-cent pitch change, lowering the pitch is more frequently discriminated, while with 30-cent pitch change more often is the perception of highering pitch. The explanation of this phenomenon can be given against the background of categorical perception. When the test task comprised two tones of the same pitch, the situation corresponded to the occurrence of quasi melodic prime. Such interval could be produced even when the second tone in the test task was changed by 15 or 30 cents upward or downward. It should be mentioned that 30-cent pitch change approximates the value of causing the break up of the category (zone) of prime, that, according to Garbuzov (1948) is 24 cents. The research made by Rakowski (1978), devoted to recognizing interval categories, established that the hypothetical border between the category of prime and the category of minor second is 30 cents. In accordance with the character of this interval, a stronger impression of its stability is made by a slight diminishing of the second tone (Fyk, 1980) and this can be the cause of better detection of upward 30-cent pitch change than downward one.

Changing the pitch of the second tone by 75 cents upward or downward led, in fact, to the production of a new category of interval, i.e. minor second. It seems that the interval produced through changing the tone pitch by 75 cents downward assumes to a greater extent the character of minor second than the interval produced through identical pitch change upward. Commenting upon the fact, that in the present study the pitch change direction affected the discrimination of 75- and 30-cent pitch change, may lead to the conclusion that such pitch changes - consistently with the categorical perception of tone pitch - were perceived by children as pitch changes lying on the border of categories, therefore they can be referred to as between or intercategorical. For children they had different perception values than 60- and 45-cent pitch changes contained within the category which due to their character, can be called within or intracategorical.

Acknowledgements

This research was supported by The State Committee for Scientific Research, Grant No 1 P107 022 07.

References

BLAGONADIERZHYNA, L.V. (1940). Psychological analysis of the perception of melody.


The McGill Piano Project: Effects of piano instruction on children's cognitive abilities

Eugenia Costa-Giomi
McGill University

It has been suggested that certain musical processes might involve regions of the brain responsible for the processing of visual-spatial information (Zatorre, 1994). Several studies have explored the relationship between musical and spatial abilities (Barret & Barker, 1973; Karma, 1982; Maniurzewska, 1978; Hassler, Birbaumer, and Feil, 1985; Hassler & Nieszlag) and musical training and spatial abilities (Costa-Giomi, 1996; Hassler, 1992; Rauscher et al, 1994; Hurwitz et al, 1975). They have generally found that musically talented or musically trained individuals tend to obtain high scores in tests of spatial abilities. Most of these studies computed correlations between the performance in music and spatial tasks or compared the spatial scores of musicians and nonmusicians to explain the relationship between spatial and musical abilities. The only study which investigated this relationship by actually implementing a treatment (music instruction) with a small number of children suggested a causal relationship between the two types of abilities (Rauscher et al, 1994).

The purpose of the longitudinal study was to investigate the effects of piano instruction on selected academic and nonacademic behaviors of children. This report will present the results concerning children's cognitive abilities.

Methodology

Sample

A letter describing the project was sent to the parents (n= 698) of all fourth-grade children (9-year old) attending the 20 English schools of the largest school board in Montreal. Two hundred and eighty-nine parents stated interest in the project. Of the parents who did not respond to the letter, 43 were contacted by phone and 74 by mail. All interested parents were sent a questionnaire on the basis of which the selection of the sample was made.

The 117 children selected for the project had never participated in formal music instruction, did not have a piano at home, and their family income was below $40,000 CDN. An effort was made to include all single-parent families interested in the project (30%), and of diverse ethnic and educational backgrounds. Girls and boys were equally represented in the sample. Approximately 25% of the children had unemployed parents. Of these families, 97% reported welfare subsidies of less than $20,000.

Sixteen of the original 20 schools were represented in the sample. Sixty-three children from 10 schools were assigned to the experimental group (group receiving piano lessons) and 54 children from 13 schools to the control group (group not receiving piano lessons). Because of practical problems (e.g., school schedules) and ethical concerns (e.g., parents and children feeling discriminated for not being provided with piano lessons), three schools had no control children and four schools had no experimental children. The sample from the school attended by 10 children who dropped out of the project was not included in the analyses because the preservation of the randomness of the remaining sample (n = 10) was questionable. Sixteen children were absent on a day in which the tests were administered; incomplete data from these children were not analyzed. The final
sample was comprised of 81 children, 37 in the control group and 44 in the experimental group.

Treatment

Each child in the experimental group received two-years of piano instruction and an acoustic piano at no cost to the families. Pianos were delivered to the four schools that did not have adequate pianos and to the homes of the experimental children.

Nine teachers (six female and three male) of the McGill Conservatory were selected to participate in the project on the basis of their experience in teaching piano to children. Children received individual piano lessons weekly. The 30-minute lessons were scheduled during lunch or after-school hours at the participating schools.

Teachers followed the curriculum guidelines of the McGill Conservatory. The guidelines were flexible to allow the teachers to select repertoire and teaching methods appropriate to the individual students.

Testing

Children in both the control and experimental groups were administered five standardized tests before the beginning of the treatment: Coopersmith Self-Esteem Inventories, Developing Cognitive Abilities Test, the language and mathematics subtests of the Canadian Achievement Test, the fine motor subtests of the Bruininks-Oseretsky Test of Motor Proficiency, and the tonal and rhythmic audiation subtests of the Musical Aptitude Profile. Tests were administered in groups during school hours. The only individual test was the motor skills test. At the end of the first and second year of instruction, all children were administered the self-esteem and cognitive abilities test. At the end of the second year of treatment children also took the academic achievement test.

The Developing Cognitive Abilities Test provides four scores, verbal, quantitative, spatial, and total. Children took the level "E" of the test at the beginning and end of the first year of the project and level "F" at the end of the second year.

According to the results of independent t tests, there were no differences between the scores of the control and experimental groups in any of the five standardized measures (cognitive abilities, academic achievement, self-esteem, motor skills, musical abilities) administered prior the beginning of the treatment.

Results

Only the results related to the cognitive abilities test were analyzed for the present report. An analysis of variance (ANOVA) with repeated measures was performed for Group (experimental or control) and Income (<$20,000 $20,000 - $30,000, $30,000 - $40,000) on the total scores of the cognitive abilities test in 1994, 1995, and 1996 (repeated measure: Year). The Group * Year interaction was found to be significant ($F[2,150] = 3.56, p = .03$). Analyses of simple effects showed no difference between the scores of the two groups of children in 1994, but significant differences between those in 1995 ($F[1,75] = 8.30, p = .005$) and 1996 ($F[1,75] = 13.19, p = .001$). These analyses also indicated that the scores of the control group changed significantly from year to year ($F[2, 150] = 37.97, p < .001$) and so did those of the experimental group ($F[2, 150] = 53.62, p < .001$).
Figure 1 shows that the experimental group obtained higher scores than the control group in 1995 and 1996.

In order to study the differences between the control and experimental groups further, three ANOVAs with repeated measures were performed on the three subtests of the cognitive abilities test: verbal, quantitative, and spatial. As in the previous analysis, Group and Income were the independent variables and Year, the dependent variable (repeated measure). In the analysis of each of the subtests, Year was found to be a significant main factor; children's scores changed significantly from year to year.

A significant Group * Year interaction was found for the spatial subtest scores \( F[2, 150] = 8.21, p < .001 \). Analyses of simple effects indicated similar results to those of the total scores in the cognitive abilities test. While no differences between the 1994 spatial scores of the control and experimental groups were found, significant differences between the groups were established for the 1995 \( F[1, 75] = 8.59, p = .004 \), and 1996 \( F[1, 75] = 9.87, p = .002 \) scores. Figure 2 shows that the experimental group scored higher in 1995 and 1996 in the spatial abilities subtest than did the control group.

A significant Income * Year interaction was found for the quantitative subtest scores \( F[2, 150] = 2.59, p = .04 \). Analyses of simple effects revealed that the scores changed significantly from year to year for each of the three income groups. No significant interactions were found in the analysis of verbal scores.

Discussion

The results of this study suggest that piano instruction affects the cognitive abilities, especially spatial abilities, of children. The children who participated in two years of piano instruction scored significantly higher in the cognitive abilities test than those who did not receive any formal music instruction. While no differences in pretest scores could be established between the groups, significant differences were found for their scores after one and two years of treatment.
The analyses of the three subtests of the cognitive abilities measure, verbal, quantitative, and spatial subtests, indicated that piano instruction affected children's spatial abilities. No differences in spatial abilities between the groups could be established prior to the beginning of the treatment. However, children receiving lessons scored significantly higher in the spatial subtest than those not participating in piano instruction after one and two years of treatment. These results provide support to the notion that there are common processes involved in the performance of musical and spatial tasks and that the development of the ability to perform the former develops the ability to perform the latter.

Piano instruction did not affect children's verbal abilities. However, it had an intriguing, though not significant, effect on their quantitative abilities. The study of the Income * Group * Year interaction (p = .09) suggests that piano instruction seemed to affect children's quantitative ability differently depending on their family income: the higher the income of the families the more positive the effect of piano instruction. This is in agreement with previous findings of this longitudinal project which showed that one year of piano instruction improved the quantitative abilities of children with family incomes between $20,000 and $40,000 but not those of children with lower incomes (Costa-Giomi, 1996). Given the reduced income brackets used in this study, further research on the effects of music instruction on quantitative abilities with populations with higher incomes would be desirable.

It is hoped that future neurological studies and longitudinal investigations on the effects of musical training on cognitive abilities will increase the understanding of the cognitive processes involved in musical tasks. This knowledge would help educators and psychologists optimize children's educational opportunities to develop their capabilities.

This project was supported by a grant from The National Piano Foundation

References


Intellectual activity in the musical domain can be defined as an ability to combine implicit or explicit auditory representations with musical operations. From the point of view of cognitive development, the most important question concerning music compositional abilities of children is how complex musical structures children at different ages are able to understand.

Piaget’s reliance on logical structures has been criticized for not fitting well in the musical domain, nor doing justice for children’s thinking. The theory of Robbie Case (1985) is used in this paper as an alternative framework within which the empirical results of developmental research in music could be understood. Case’s idea of the hierarchical integration of executive control structures is particularly applicable to productive use of musical knowledge, because an executive control structure includes the representation of the child’s current state, the representation of some other state the child voluntarily pursues and the representation of a sequence of operations that will take the child from the current state to the more desirable state (Case 1985, 412). The theory is applied to form hypothesis of musical control structures of children of 0-11 years.

Problems with Piaget’s theory

Till our day the dominating theory of cognitive development has been the theory of Jean Piaget, who looked at the developing child from the point of view of logical structures. He also derived some of his developmental hypothesis directly from logical structures (Flavel 1963, 188-195). For example, the concept of conservation is based on logical reversibility, involving compensation. However, the structures of musical knowledge are not necessarily logical. While compensation is possible e.g. within the temporal dimension (see Hildebrandt 1987, 86), it is impossible for the spatial and temporal dimensions to compensate each other. There is no such thing as 'conservation of melody', because syntactical dimensions (melody, harmony and rhythm) which are spatiotemporal dimensions - and important in music composition - cannot compensate each other.

It has also been discovered that a child mastering some logical operation with one kind of material often fails with another kind of material (see Markman 1989), because contextual factors have effect upon the focus of attention. The dynamics of development cannot be understood without considering attentional and short-term processes.

The theory of Robbie Case

The theory of Robbie Case (1985, 1992) suggests that learning is task-specific, domain-specific and constrained by the short-term storage space and attention. Common to the development of different domains is the form of hierarchical integration. Cognitive development is divided in four major stages: the stages of sensorimotor (4-18 months), relational (1.5-5 years), dimensional (5-11 years), and vectorial (11-18 years) operations. Case’s basic concept is called an executive control structure. It includes A) a representation of a current state, B) a representation of a more desirable state and C) a mental model of an action leading from one state to another. Executive control structures are constructed in specific learning experiences. Transition from one stage to the next is brought about by hierarchical integration of executive control structures constructed at the previous stage. Integration is due to automatization, which Case calls operational consolidation. Within a stage there are substages of 1) unifocal coordination, when a new structure can be applied in isolation, 2) bifocal coordination, when two such units can be applied in succession, 3) elaborated coordination, when two or more units are applied simultaneously and integrated into a coherent system. As a result of coordination operational efficiency increases. This means that the short-term storage space
grows. Operational efficiency depends also on the biological maturation of the brain (Case 1985, 421-414; Case 1992, 365-366.)

**Memory processes, executive functions and music composition**

Children’s and adults’ processing of musical knowledge and development of musical skills are somewhat parallel. Similarities can be found when comparing developmental phases of an adult improvisational skill and children’s knowledge structures. According to Clarke (1988), at the level when improvisational skill has not fully developed associative and repertoire-based thinking is more common than hierarchical thinking. Similarly, preschool children do not usually modify their productions extensively. Generally they produce associative chains consisting of pairs of musical patterns, and little later on form versions based on a larger
repertoire of patterns (Dowling 1982). In the age of 7-11 there is an increasing tendency to modify and develop motives, and the ability to remember the whole product develops (Kratus 1985, 1989). In this sense, the ability to compose develops particularly in middle childhood.

What develops in musical knowledge is the hierarchical differentiation and integration of pitch-time-units, and flexible use of different levels. The degree of organization is reflected in improvisational/compositional process, and to the structure of its products. The process of both adults and children at various ages is influenced by the constraints of short-term storage space with regard to the temporal length and complexity of the event.

Developmental neuropsychology has revealed that executive functions developing during childhood are related not only to myelinization (as Case assumed 1985) but to maturation of the frontal cortex, which is an integrating organizer of connection systems between different regions in the brain, and is activated in nonautomatic, high priority actions (Thatcher, 1991). Prefrontal lobes display a multistage process of development. Rudimentary skills as recognition develop very early, but more complex skills like planning continue to develop during adolescence (Welsh et. al 1991). Composition involving modifying, testing and repeating musical material could perhaps be compared to the planning factor of executive function. So, an adult novice is able to use rational planning and metacognitive strategies in her/his learning process, while the small child lacks this kind of top-down skills.

The direct result of this is that in the beginning of development a child controls limited time-spans in her/his improvisation or composition. Musical models that get stored in the long-term memory are temporally limited fragments. What is in the focus of attention of the child are the relations within that limited musical sequence. The solutions of the child tend to reflect probabilities of relatively local musical levels. Part of the solutions reflect probabilities of action: when a good procedure is discovered it is used more often.

It is also noteworthy that more simple modes of executive function do not disappear during the development, but become embodied components of the more complex functions (see Cowan 1995). When sensorimotor control of music production is inhibited, it is obvious that results in music production get poorer. E.g., 7-year-old children who had an opportunity to use 'the mouse' for graphic modifications, but not to play or sing the music, produced pieces that lacked melodic contour and figural grouping (Wilson & Wales 1995).

Musical development and the theory of Case

An application of Case’s mechanism to the musical domain could be formed as follows:

![Developmental stages of music from sensorimotor to dimensional.](image)
At the sensorimotor stage the infant controls changes in the basic parameters of music. As a result an early form of musical pattern emerges. At the relational stage the child controls polar relations between patterns (repetition, alteration) and polar relations within patterns (melodic contour, intervallic relations, relative durations, pulse, tempo, relative tonal stability). As a result of integration of relational control structures deeper levels of hierarchies begin to emerge. At the dimensional stage the child becomes able to control conflicting effects on different levels of hierarchies, without losing the whole. Attention can be focused on non-phase conditions of grouping and meter, interval/key relations, relative/abstract tonal stability, local/global tonal stability and horizontal/vertical dimensions of tonal harmony.

Musical control structures at the sensorimotor stage

At the sensorimotor stage sensory representations and gross-motor schemes get co-ordinated. In premusical production infants produce sounds by muscular tension in the larynx or by moving limbs, and monitor the results in the sound. It could be assumed that infants control tension and rest with three basic parameters of pitch, time and intensity using the periferal input of the muscles.

The auditory abilities of infants are remarkable in many ways. Infants discriminate single pitches and melodies on the basis of contour; the discrimination of changes is better in diatonic than non-diatonic contexts (Chang & Trehub 1977; Trehub 1994). They also segment musical continuum using the same principles than adults (Krumhansl 1990; see Lerdahl & Jackendoff 1983). They cannot coordinate these auditory representations to their sound production schemes very well, though.

At the stage of operational consolidation sound production is not differentiated of other activities. At the stage of unifocal coordination sound production and sensory representation of sound get co-ordinated, as well as movement and sound. Direct imitation of vocal or more rhythmic sounds that have occurred previously in the vocal play is now possible. E.g. in the age of 3-6 months infants start to produce movements and gestures actively when they hear music (Moog 1976). It is assumed that the infant is able to focus attention to some parametrical change in her vocal play. At the stage of bifocal coordination reciprocal communication and monitoring the results in a more specific way develops. The infant is now able to control a parametrical change that is analogous to changes in familiar premusical sounds. At the stage of elaborated coordination the infant controls several parametrical changes. Imitation of new sounds is possible. In the end of the stage vocal glissandi and rhythmic babbling get integrated and form a song with discrete pitches (see e.g. Moog 1976; Moorehead & Pond 1978). An early form of a song pattern emerges. Melodic contour is, however, unidirectional and rhythm is bound to syllable repetition. Each repetitive act is regarded as a sensorimotor operation.

Musical control structures at the relational stage

Spontaneous production of a musical pattern is a complex sensorimotor control structure. When regarded as a relational control structure it can be apprehended as a musical unit. As in language there are relations between units and relations within units. The possible relations between musical patterns are repetition and alteration. Possible relations within a musical pattern are the direction of melodic contour, relative duration (long-short), relative tempo (fast-slow), regular speed of release of two events (pulse), the direction of a melodic interval, the relation of two simultaneous pitches, the relation of a pitch and the reference pitch of the context, the relation of the pitch and the octave, and relative stability/instability (see Krumhansl 1990).

At the stage of unifocal coordination the child focuses on a relation between two patterns or one relation within a pattern. The song forms a long chain, where the next event is determined by maybe only one previous event. At the stage of bifocal coordination the child coordinates the operations of relating two units and relating within a unit. The child can e.g. focus on contour direction and on intervals, or on temporal relations in the pattern. Stable relations of pitches within a pattern are gradually formed (see Davidson 1994). If there is a reference pitch, it is lost.
across the combinations. In temporal relations a more complex grouping that synchronizes
durations with a regular tempo of release develops within the pattern. All of these
possible relations of pitch and time are not mastered in the same pattern, or across patterns.
**Elaborated coordination**: Previously mastered coordinations of pitch and time within and
between units get integrated into a complex relational control structure. This results a
continuous coherent song production. The key context, however, is 'flat'. It is only in the end
of the substage when the child is capable to focus attention on both the whole and the parts.
Therefore thinking can be characterized as bound to the horizontal continuum.

Other important relational control structures are the relation between words and music and the
use of symbolic tools. Using instruments and graphic notation sets extra demands to the child's
information processing resources. Vocal improvisation is mastered best.

**Musical control structures at the dimensional stage**

At the dimensional stage the linear whole can be broken and reorganized at the level of a single
surface event. Because of this differentiation more subtle problems arise: conflicts between the
surface level and deeper levels. The child is assumed to become able to control hierarchic
relations in music. At the end of the stage the child is able to focus attention on the conflicting
effects of different hierarchical levels in music compositional tasks. It is unclear, however, how
much practice is needed.

Conflicts in temporal relations exist between grouping and meter (see Bamberger 1991; Upitis
1987). At the beginning of the stage the child is able to focus on either grouping or meter. Little
later on the child can focus on both of them if they are in phase. Thus the composition at the
dimensional stage could eventually include voluntary integrated control of non-phase conditions
of rhythm (see Lerdahl & Jackendoff 1983).

At the dimension of pitch analogous conflict could be between melodic movement (contour, the
direction of a successive interval) and the tonal scale. At the end of the stage the child is assumed
to control both the melodic operations (transposition etc.) and relations within the scale. Also
the tonal functions begin to emerge (Krumhansl & Keil 1982). Conflicts probably arise
between local tonal tensions in groups and a more global tonal movement to tension and return
to the tonic. This requires the integration of tonal stability hierarchy and the temporal hierarchy
(see Imberty 1981), and thus represents elaborated level of production.

Local tonal harmony can be understood also on the basis of successive (=melodic, horizontal)
and simultaneous (=chord, vertical) intervals. The simultaneous and successive dimensions
can be in conflict. It is assumed that a child performing a compositional task focuses on one of
these dimensions in the beginning of the dimensional stage, and both at the end of the stage.
This kind of harmonic control does not necessarily mean that stability hierarchy would be fully
developed.

**References**

Musical Intelligence. The United States of America: Harvard University Press.
Case, Robbie 1985. Intellectual Development. Birth to Adulthood. The United States of
America: Academic Press.
Case, Robbie 1992. The Mind’s Staircase. Exploring the Conceptual Underpinnings of
Children’s Thought and Knowledge. The United States of America: Lawrence Erlbaum
Associates.
Chang, H. & Trehub, S.E. 1977. Auditory processing of relational information by
Generative processes in music. The psychology of performance, improvisation and

455


Cross-Cultural Perspectives On Young Children's Music-Making
In The Context Of Play
Danette Littleton
The University of Tennessee at Chattanooga - USA

Play in human history is the cradle of art. Huizinga (1938), in his classic study on *homo ludens*, man the player, stated that humans in their earliest stages played, and that the "play spirit" of joy, pretend, and non-seriousness has been a major civilizing force in human development. Throughout history, according to Huizinga, play and music were bonded by values that transcended logic, the visible, and the tangible. Huizinga explained that the primary quality of play is aesthetic quality. He proposed that the play-spirit reached its full flower in western civilization in the 18th-century and attributed the perfect balance of its play-content and its aesthetic content to the "supreme importance of 18th-century music." For "music," Huizinga declared, is the "highest and purest expression of the *facultas ludendi*" (p.187).

Play Theories

Play and Child Development

Play and Children's Musical Development
With the notable exception of Moorhead and Pond (1978), few researchers have demonstrated interest in studying children's play regarding music and music settings. Further, when a "free play setting" is the focus of investigation, the term, "free play" may be inappropriately exchanged with "naturalistic setting" (Miller, 1987), or misused to describe "teacher guided play" (Swanwick, 1986, Metz, 1989) or substituted for "playfulness" (More, 1987). Gunsberg (1982) developed "play training" techniques and named the results "improvised musical play." Without a clear definition of what play is and is not, understanding of children's play with music in the context of development remains limited.

Play provides the ideal medium for observational investigations of children's music behaviors set apart from direct adult instruction, suggestion, or guidance. Studies of young children as they play freely in a music-specific environment reveal music experiences uniquely child-made and distinctly unlike those directed by adults (Moorhead and Pond, 1978, Littleton, 1991). Moorhead and Pond examined the spontaneous music-making of children aged two to six at the Pillsbury School, Santa Barbara, California from 1937-1948. Conclusions about children's social and musical behaviors in free-play music settings were summarized:

1. A natural play environment with musical instruments stimulates
the young child's musical imagination through free experimentation with instrumental sounds.

2. Creative song in young children emerges with speech and chants through spontaneous play.

3. Socially, young children learn to collaborate spontaneously in musical activities with a minimum of adult intervention.

4. Creative rhythmic movement occurs simultaneously with children's instrumental and vocal musical productions.

Following the Pillsbury Studies, interest in pedagogy and study of children's free-play with music declined in the USA. Attention to children's musical development within the larger framework of cognitive development dominated classroom practice and research inquiry. From the 1950's, preference for adult-directed, group, music activities persisted in preschool classrooms where children learned by rote to perform songs and move rhythmically to music selected by an adult. Music, as a mnemonic device to teach basic skills, colors, numbers, spatial direction, and subject content prevailed in many preschool settings. Children's music learning environments lacked time for spontaneous music-making, space for music-play to occur, and free access to musical instruments.

Interested in children's non-directed, spontaneous, music-making, Littleton (1979, 1991) studied 3-, 4-, and 5-year-old children's cognitive, social, and music behaviors in the context of play. Results of those studies prompted this investigation of cross-cultural comparisons of children's social and music play behaviors when time, space, and access to musical instruments are provided.

The present study compared spontaneous music behaviors of 4-, and 5-year-old children playing freely in a music-specific play setting in Nishinomiya City, Japan with children similarly grouped in a similarly prepared music play setting in Chattanooga, Tennessee, USA. Unobtrusive, systematic observations were made on videotape and data collected to answer the following questions:

1. What musical behaviors do 4-, and 5-year-old children in Japan versus 4-, and 5-year-old children in the USA exhibit in a free play, music-specific setting according to: a) spontaneous song or chant, b) spontaneous improvisations with musical instruments, and c) spontaneous movement to music?

2. What social play behaviors do 4-, and 5-year-old children in Japan versus 4-, and 5-year-old children in the USA exhibit in a free-play, music-specific setting according to: a) solitary play, b) parallel play, and c) group play?

3. What, if any, gender differences occur as 4-, and 5-year-old children in Japan versus 4-, and 5-year-old children in the USA play freely in a music-specific setting?

Subjects, Settings, Materials

Twenty-four 4-, and 5-year-old children who comprised one kindergarten class on the campus of Seiwa College, Nishinomiya City, Japan were observed as they freely played in a music-specific play setting. Observations by unobtrusive video camera were collected in 8, 20-minute sessions with 2 alternating groups of children: group A (12 children) videotaped at 1st,
3rd, 5th, and 7th sessions; group B (12 children) videotaped at 2nd, 4th, 6th, and 8th sessions.

A music-specific play setting was equipped with musical instruments: guitar, electronic keyboards, xylophones, metalophones, marimba, tone bells, triangles, jingle bells, small and large symbols, small and large drums, castanets, wood blocks, maracas, guiros, and equipment, including a music stand, music books, and a conductor's baton.

Twenty-two 4- and 5-year-old children who comprised one kindergarten class at St. Nicholas Episcopal Day School, Chattanooga, Tennessee, USA were observed as they freely played in a music-specific play setting. Observations by unobtrusive video camera were collected in 8, 20-minute sessions with 2 alternating groups of children: group A (11 children) videotaped at 1st, 3rd, 5th, and 7th sessions; group B (11 children) videotaped at 2nd, 4th, 6th, and 8th sessions.

A music-specific play setting was equipped with musical instruments: guitar, electronic keyboards, xylophones, metalophones, piano, violin, tone bells, triangles, jingle bells, small cymbals and Chinese gong, small and large drums, wood blocks, maracas, guiro, and equipment, including a music stand, music books, and a conductor's baton.

Design, Data Collection, and Analysis

Quantitative and qualitative methods of inquiry were used to generate information needed to answer the research questions in this study. Time-sampling procedures provided frequency counts of selected music play behaviors, social play behaviors, and gender. Anecdotal recording of the children's experiences provided descriptive data to augment and enhance the interpretability of the quantitative data. All behaviors for description and analysis were recorded by a non-obtrusive video-camera for scoring, examination, and analysis on later play-back.

Upon arriving at the music play settings, children were told that they would be videotaped as they played. A minimum of two adults were present during the observations: the researcher, operating the video camera, and a preschool teacher who served to monitor non-musical behavior as appropriate. A minimum of adult intervention was maintained.

Coding instruments were designed by the researcher for time-sampling and anecdotal recording. At the end of a 1-minute interval, the targeted behavior of an individual child selected a random was observed and recorded. At the end of each subsequent minute, the other children were observed individually in a random sequence, and their behavior recorded until all children in the setting had been observed. Observation and recording of behavior continued in the prescribed sequence until 24 observations in the Japan study and 22 observations in the USA study were completed. Frequency counts of selected play behaviors, and gender provided numerical data for comparisons.

Collection of anecdotal data was performed by reviewing each tape and making transcriptions of dialogue and music play events. Narrative descriptions in the form of summary notations for each 20-minute tape focused on noteworthy music and social play behaviors within the chosen content limits of the study.

The system of time-sampling was tested for reliability by determining a significant degree of consistency of observations between the researcher and coders. The coders were trained to memorize categorical definitions of the targeted music and social play behaviors. Using the coding instruments, observations were made of each play category and gender for 2 complete 20-minute tapes. Upon completion, the coding sheets were checked, and a 98% inter-observer rate of agreement was obtained.
Results

Analysis of time-sampling data revealed differences across cultures in the social play of these 2 groups of children: a) more solitary play among Japanese children than USA children; b) more instances of parallel and group play among USA children than Japanese children; c) fewer instances of non-play by Japanese children than USA children.

Analysis of time-sampling data revealed differences across cultures in the music play of these 2 groups of children: d) more instances of vocal, and movement music behaviors among USA children than Japanese children, e) more instances of music instrument playing behaviors among Japanese children than USA children, f) similar music play behaviors of boys in Japan and USA, and g) similar music play behaviors of girls in Japan and the USA.

Anecdotal data revealed less dialogue during play and fewer instances of fantasy play among Japanese children than USA children. In each setting, the presence, attraction, and holding power of selected musical instruments disclosed the children's capability for sustained interest in music-making in a music-specific, free-play setting.

Conclusions

Additional research is needed to examine the complex issues of young children's spontaneous music-making across cultures. Expanded knowledge and understanding of: a) the aesthetic dimension of children's play with music, b) the play-content of children's music-making, c) stage theory applied to children's play with music: infant, toddler, preschool, and school-age, d) cognitive and social benefits of children's play with music, e) gender differences in the music-play of girls versus boys, f) structured versus non-structured music learning environments for children, namely, teacher-directed large and small group activities, teacher-guided large and small group activities, and child-initiated, free-play activities, g) attraction and holding power of music instruments, h) indoor and outdoor music play spaces, i) children's music play behaviors in the home versus the classroom setting, j) the play origins of musical creativity, k) creative music processes of young children versus adults, l) researcher-initiated versus child-initiated tasks in the study of musical creativity, improvisation, composition.

How individual and groups of children across-cultures play with music is worthy of investigation for what can be learned about the play origins of musical creativity and its prominence in human development.
References


Music Performance: Measurements
Timing variations in music performance: hierarchical segmentation organisation and rhythmic grouping
Amandine Penel and Carolyn Drake
Laboratoire de Psychologie Expérimentale, 28 rue Serpente, 75006 Paris, France

INTRODUCTION
Traditionally, systematic performance variations have been attributed to the performer's wish to transmit to the listener various aspects of the musical structure (Clarke, 1988; Gabrielsson, 1987; Palmer, 1989; Repp, 1992), such as phrase, metric, rhythmic or voice structure. These performance variations refer to relatively high level processes involving a global analysis of information which requires the use of musical knowledge. Here, we studied temporal variations, in particular the duration between the onsets of successive events (inter-onset intervals - IOI). We suggest that, in addition to these variations produced by relatively high level processes, other systematic temporal variations would result from lower level processes involving a local analysis of acoustic surface features of the music, in particular the grouping of several notes into perceptual units (Deliege, 1987): some events would be perceived shorter and consequently played longer as a phenomenon of perceptual compensation (Drake, 1993a). This is similar to the "bottom-up" hypothesis proposed by Repp (1995). Several sources of evidence in favour of some performance variations being related to lower level processes emphasise their universal and unavoidable nature. The comparison of the reproduction of musical rhythms by groups of subjects varying considerably in levels of passive musical acculturation (age) and musical expertise (training) indicated that the same systematic temporal variations related to rhythmic groups were observed for 5-year-old children, adult non musicians and musicians. More precisely, in a production task, in a series of tones separated by short intervals, the last interval was systematically lengthened by between 10% and 20% in relation to the preceding interval (Drake, 1990; 1993b). A series of experiments (e.g. Drake, 1993a) using various perceptual tasks (temporal adjustment, temporal irregularity detection and identification) rejected the possibility that this variation could have been produced by a motor factor. This variation, which will be referred in the following as a variation in relation to rhythmic groups, seems also unavoidable: when pianists were asked to perform simple musical sequences in either a musical or a mechanical fashion, this variation hardly decreased in the mechanical performances, contrary to other patterns of variations (Drake & Palmer, 1993). Converging support for variations produced by lower-level processes can be found in a series of studies by Repp (e.g. Repp, 1992; 1995) which suggest that, when played mechanically, some intervals are perceived shorter (or longer) than they are played, which makes, for the listener, a lengthening difficult (or easy) and a shortening easy (or difficult) to detect, and which results, for the performer, in a lengthening (or a shortening) of these.

The first aim of this study was to show that in the same normal musical context, performers produce variations to highlight some aspects of the musical structure (as a result of high level processes), and also variations as perceptual compensations (as a result of lower level processes). To demonstrate this, we analysed global variations (above the note level) as a possible reflection of hierarchical segmentation organisation, i.e. the hierarchical segmentation of the music into phrases (a high level process), and local variations (at the note level) as a possible reflection of the segmentation of the musical surface into basic rhythmic groups (a lower level process). Global variations reflecting hierarchical segmentation organisation were analysed in relation to Todd's 1985 model of phrase-final-lengthening (Todd, 1985). Local variations reflecting the segmentation of the musical surface into basic rhythmic groups were analysed in relation to the results of Drake and Palmer's 1993 study. The second aim of this
study was to evaluate to what extent temporal variations could be identified. Regression analyses gave an indication of the part of variability explained and allowed a comparison of the relative importance of those two sources of variations. Finally, we re-examined these findings in the light of a psychological segmentation model in order to assess whether a process-oriented approach may present a harmonious model of performance variations.

METHOD

Eight professional pianists performed Schumann's Träumerei three times with the same interpretation. We analysed the first eight measures of the piece, with the repeat. Recordings were carried out on the Yamaha Disklavier II, monitored by a computer. Using the "expressive timing" function of POCO (Honing, 1990), text files were obtained for each performance, with the onset of each note of the soprano voice. We defined the relative temporal variation as being the normalised ratio of the duration of the IOI as it was played in relation to the duration of this IOI in the score.

RESULTS

Intra-individual and inter-individual correlations were computed. They were all significant (p < .001), with a mean correlation of \( r = 0.80 \), ranging from \( r = 0.46 \) to 0.97. Thus, for each pianist, timing profiles were consistent across the different performances and repetitions and the profiles of different pianists had many characteristics in common. To obtain an average timing profile for our group of subjects where non-systematic variations would be reduced, an arithmetic mean was calculated over the eight individual mean timing profiles (see Figure 1).

![Fig. 1: Average timing profiles for the two repeats (three performances by eight pianists).](image)

Analysis of global variations

Global variations reflecting hierarchical segmentation organisation were analysed in relation to Todd's 1985 model of phrase-final-lengthening which was taken as an indication of the segmentation into phrases. Figure 2 presents the predicted and observed variations at the measure level.
Phrase-final-lengthening can be observed. As the model was not designed to account for the accelerands observed at the beginnings of the phrases, it is not surprising to observe differences between predicted and observed variations. To compare the lengthenings within each phrase, we analysed the mean duration of the last measure in each phrase. An ANOVA on the mean measure duration by repeat (2) and phrase (2) revealed no significant effect of repeat, a significant effect of phrase ($F(1, 23) = 228.8, p < .0001$) and a significant interaction between the two ($F(1, 23) = 9.66, p < .005$). Thus, phrase-final-lengthening was greater in the second phrase of each repeat than in the first phrase, and this effect was stronger in the second repeat. Pianists therefore segmented the excerpt between phrases, between repeats, and at the end of the second repeat, thus providing evidence for three hierarchical levels of segmentation.

Analysis of local variations
Local variations reflecting the segmentation of the musical surface into basic rhythmic groups were analysed in relation to the results of Drake and Palmer's 1993 study. In this study, the last interval of a rhythmic group was played, on average, 10% longer than the preceding one. Figure 3 presents the four rhythmic groups of more than two notes in the excerpt and Figure 4 the variations observed for these rhythmic groups.

Fig. 3: Rhythmic groups of more than two notes in the excerpt.

Fig. 4: Variations in relation to the segmentation into rhythmic groups. The relative temporal variation of each note of each rhythmic group is plotted against its position in the rhythmic group. The last, long note is in position 0, the second to last in position -1, etc. Thus, the note in position -1 is predicted to be longer than the note in position -2. The amplitude of the observed lengthening is indicated between parenthesis.
Variations corresponding to the segmentation into rhythmic groups can be observed. An ANOVA on the relative temporal variations by group (4) and position in the groups (the second to last and last interval) confirmed a significant effect of position in the group \( (F(1, 23)=239.8, p<.0001) \), with the last interval (-1) longer than the preceding one (-2). There was also a significant effect of group \( (F(3, 69)=41.6, p<.0001) \) and interaction between the two factors \( (F(3, 69)=62.5, p<.0001) \), indicating that the lengthening was greater for groups 1 and 3 than for groups 2 and 4. This was due to the presence of arpeggios on the left hand on the last notes of these groups.

**Contribution of these variations to overall variability**

The analysis of global and local variations showed that some variations reflect hierarchical segmentation organisation and that others reflect rhythmic grouping. How much variability can be explained by these sources of variation? And what is the relative contribution of these two sources? To address these questions, a forward stepwise regression analysis was conducted, taking the average timing profile as the dependent variable and variations corresponding to the two possible sources (hierarchical segmentation organisation and rhythmic grouping) as the independent variables (now using absolute values of intervals reduced to the equivalent eighth note for long notes). This analysis was first carried out using variation values predicted in the literature in order to establish the percentage of variance explained by these values. Then, the same analysis was carried out using variation values obtained in the previous analysis. The following regression equations and percentages of variance explained were found (\( F \) to enter = 4; dependent variable = average timing profile; independent variables = deviations from minimum tempo of global variations corresponding to hierarchical segmentation organisation and of local variations corresponding to rhythmic grouping):

Variation values predicted in the literature: Total:

\[
Y = 514.13 + 2.62 \times \text{(rhythmic groups)} + 0.99 \times \text{(hierarchical segmentation)} \text{ ms.}
\]

17.6% 14.3% 31.9%

Variation values obtained in the analysis:

\[
Y = 436.34 + 1.20 \times \text{(hierarchical segmentation)} + 0.77 \times \text{(rhythmic groups)} \text{ ms.}
\]

33.1% 24.9% 58.0%

(The independent variables are in order of entry in the regression).

Thus, in both cases, a significant part of variance was explained by variations related to the hierarchical segmentation organisation and by variations related to the segmentation into rhythmic groups. Moreover, in both cases, the size of the contributions was similar for both organisational principles. As predicted, the effects were larger when the values obtained in the analysis were used.

We also examined the possible influence of melodic groups and of the metric structure. Variations were found non-systematic in relation to melodic groups. Variations were found systematic in relation to the metric structure but did not contribute significantly towards explaining the variance.

**A hierarchical segmentation model**

Until recently, the general approach to the study of music performance has been to emphasise how systematic variations are related to the musical structure. We propose a change in emphasis, whereby performance variations are studied in relation to the psychological processes allowing the musician to perceive the musical structure. Because of perceptual and cognitive limitations (including memory constraints), the musician has to reduce information in some way. We propose here a psychological segmentation model of temporal organisation and music performance. In a first step, the musical surface is segmented into basic groups. The duration of notes is one criterion for this segmentation, creating rhythmic groups. However, if one only takes into account this criterion for the segmentation into basic groups, our excerpt
would have a basic group of 13 notes (see rhythmic group 2), which is too long. Considering also motor music, other criteria are necessary to segment the musical surface. Pitch jumps and turns, as well as particular metrical positions may constitute opportunities for additional segmentations. Here, suppose the segmentation gave the probable basic groups indicated in Figure 5, one can observe that notes within these basic groups follow a profile of ritardando or a profile of accelerando followed by a ritardando (the first is only a truncature of the second). The interpolation of 2nd degree polynomials (or 1st degree when there are only two points) highlights this phenomenon. Notes longer than an eighth note were not included in this interpolation as, being already long, they do not need to be lengthened further.

In a second step, these basic units are grouped to form phrases. As notes follow within the basic groups a profile of ritardando or a profile of accelerando followed by a ritardando, basic groups follow within phrases a profile of accelerando followed by a ritardando. Figure 6 presents polynomial interpolations for the basic groups within phrases.

In a third step, these phrases are grouped to form repeats. Once again, a ritardando is observed for each repeat from the first phrase to the second. Finally, in a fourth step, the two repeats are grouped to form the part of the excerpt studied. Once again, a ritardando is observed from the first repeat to the second. We hypothesize that further groupings and associated ritardandos could be observed at the overall excerpt level. We therefore suggest that similar psychological processes are operating at multiple levels of analysis, namely those of segmentation and
grouping. Moreover, these similar processes result in the same type of performance variations (a ritardando or accelerando / ritardando profile).

CONCLUSION

The analysis of global and local variations showed evidence for variations related to hierarchical segmentation organisation, a high level process, and for variations related to the segmentation of the musical surface into rhythmic groups, a lower level process. The percentage of variation explained by rhythmic grouping is similar in magnitude to that explained by hierarchical segmentation organisation, underlining the importance of including lower level sources of variation when analysing music performance. When variation values obtained in the analysis are used, the proposed organisational principles can account for 58% of the variance. 42% of the variance still remain to be identified, probably by the identification of other potential sources of variation (which are not melodic groups or the metric structure).

We have proposed a psychological segmentation model of temporal organisation and music performance. This model is based on the segmentation of the musical surface into basic groups and on the grouping of these units into increasingly larger units. Each grouping process seems to result, in musical performance, in a ritardando or accelerando / ritardando profile. A way of testing this proposed model would be first to ask musicians to segment a musical sequence from a score or from a mechanical performance into increasingly larger units, and then to compare this segmentation with the ritardandos they produce when performing the sequence. Also, the communication of this segmentation could be tested by comparing segmentations listeners can make from mechanical performances and from real performances.

REFERENCES


Cooperation in piano duet performance
Lucy J. Appleton
Department of Psychology, University of Sheffield, UK
W. Luke Windsor
Eric F. Clarke
Department of Music, University of Sheffield, UK

Introduction
Musical performance has provided a useful domain for the study of motor skill and timing. Musicians are used to being measured (e.g. recording sessions and auditions/examinations) and perform extremely complex motor tasks. Moreover, a psychological, or at least empirical, approach to musical performance has promised much to musicology, providing tools which can analyse the moment-to-moment timing, dynamics and articulation of piano performances, and suggest cognitive explanations of performance behaviour. This paper builds on these foundations, but in a way which tackles ensemble performance, an issue which has been largely avoided or ignored in work on motor skill and musical performance alike.

One prior study that has explicitly investigated ensemble performance is reported in Shaffer (1984). Shaffer collected and analysed timing data from performances of a Beethoven piano duet, concluding that performers maintain synchronisation due to their possession of a shared mental representation of musical structure. Such a mental representation, expressed as a motor program, facilitates predictive, rather than reactive performing strategies. Shaffer admits that the contribution of visual and auditory feedback to coordinated ensemble performance may be just as important, but offers no concrete data which could be used to assess the relationship between timing, motor representation and feedback. This paper presents analyses of the timing of performances of same Beethoven piece, but using an experimental design which manipulates the visual feedback available to the two performers.

Subjects and Method
The pianist subjects were two postgraduate performance specialists in the music department at the University of Sheffield who had some previous experience of playing together, and had been given the opportunity to rehearse the Beethoven prior to the experiment. MIDI data from the performances were recorded onto a laptop computer using standard sequencer software. Subsequent analysis of timing was carried out using POCO (Honing, 1990).

Four performances of the first page of Beethoven’s 4-hand piano sonata Op. 6 were recorded on one or two Yamaha Disklavier pianos under 3 different conditions: i) the two pianists playing at one piano (as intended by Beethoven); ii) the same two pianists playing at two separate pianos, arranged such that the players were back-to-back and unable to see each other; iii) one pianist at one piano while the other pianist’s previously recorded performance was played on the other piano, repeated for each pianist.

Results
There are two aspects of timing control that are of interest in this duet study: the variability of timing within each player’s performance (which may be a measure of either error variance or expression); and the degree of synchronisation between the players, and between their hands. Note that Beethoven scores the piece such that the first player takes soprano and alto voices (primo) and the second tenor and bass (seceso). Thus the dependent variables examined in this study are: i) the mean standard deviation of the inter-onset-intervals within each voice (soprano, alto, tenor, bass) taken across each performance (within-voice variability); ii) mean asynchronies
between the onsets of every pair of voices taken across each performance. The timings of the performances were normalised to account for changes in absolute tempo such that the duration of each performance was equal to 100 seconds (an arbitrary value). Both of these will be studied at two levels of the metrical hierarchy - the bar and the beat - since the level of timing control may change between conditions.

Four three-way (one-between, two-within subjects) Anovas were performed. For the first the factors are: i) Condition (1-Piano, 2-Piano); ii) Part (S, A, T, B); iii) Metrical level (Bar, Beat), the dependent variable being within-voice variability. For the second analysis the factors are: i) Condition (1-Piano, 2-Piano); ii) Part pairs (S/A, S/T, S/B, A/T, A/B, T/B); iii) Metrical level (Bar, Beat), the dependent variable being asynchrony. For the third the factors are: i) Condition (2-Piano live, 2-Piano live-with-recorded); ii) Part (S, A, T, B); iii) Metrical level (Bar, Beat), the dependent variable being within-voice variability. For the fourth analysis the factors are: i) Condition (2-Piano live, 2-Piano live-with-recorded); ii) Part pairs (S/A, S/T, S/B, A/T, A/B, T/B); iii) Metrical level (Bar, Beat), the dependent variable being asynchrony.

**Variability Analysis**

For the comparison of one- and two-piano performances there were significant main effects of condition and metrical level: the one-piano condition shows significantly more variability than the two-piano condition; and the beat level shows greater variability than the bar level. For the comparison of two-piano live and two-piano live-with-recorded performances there were significant main effects of condition and metrical level: the live-with-recorded condition shows significantly more variability than the live condition; and once again the beat level shows greater variability than the bar level. There were no significant interactions of any kind.

**Asynchrony Analysis**

For the comparison of one- and two-piano performances there were significant main effects of condition and part pair: the one-piano condition shows significantly less asynchrony than the two-piano condition; and pairs of voices played by the same player are significantly more synchronised than pairs across players (i.e. between hand synchrony is better than between player synchrony). There was also a significant interaction between condition and metrical level, such that in the one-piano condition there was more asynchrony at the bar than at the beat, while in the two-piano condition there was more asynchrony at the beat than at the bar, as illustrated in Figure 1.
The comparisons of two-piano live and two-piano live-with-recorded performances showed a number of different effects depending upon which player was attempting to track the recording of their absent partner. There was a significant main effect of condition for the secondo player, such that the mean asynchrony was greater where playing with a recording as opposed to a real player, but this was not significant for the primo player. There was a significant main effect of part pair in both cases, such that the pair of parts played by the single live performer in the live-with-recorded condition are significantly better synchronised than voice pairings that cross between the live player and the recording. For the secondo player, there was a significant main effect of metrical level, such that asynchronies were larger at the beat level, but not for primo.

For both players there is a significant interaction between condition and part pair, which demonstrates that voice pairings that cross between the two live players, although showing greater asynchrony than voice pairings within the same player, are not nearly as badly synchronised as when they cross between a live player and the recorded version of the other player. Figure 2 illustrates this for the secondo player, playing with the live and recorded versions of the primo part. As can be seen, the asynchronies diverge considerably except for the Tenor-Bass asynchrony (within player), and especially so for the inner asynchronies (Alto-Tenor, Alto-Bass and Soprano-Tenor) which cross players. The Soprano-Alto asynchronies are not shown here since these imply a comparison which is meaningless in the two-piano recorded condition (i.e. between the two recorded voices).

Discussion
The results of this study can be summarised as follows:
• Under conditions of visual feedback, there is greater overall timing variability than in the absence of visual feedback, but significantly less asynchrony. This suggests strongly that the variability is mainly expressive (rather than error variance), since one would expect error variance to be positively, rather than negatively, correlated with asynchrony.
• The monitoring of asynchrony becomes less ‘fine-grained’ in the absence of visual feedback (as reflected in the interaction between condition and metrical level for asynchrony shown in figure 1). Presumably visual feedback, and the physical proximity of the two performers playing on one piano, allows them to keep a detailed track of their relative coordination.
• Synchronisation between voices that correspond to the hands of one player is better than any voice pairing that involves another player (or a recording of that player).
• Synchronisation between voices that correspond to the hands of two live performers is rather better than that between any voice pairing that involves a live player and a recording.

Conclusions
This study clearly demonstrates that visual feedback contributes significantly to both the accuracy and expressive freedom of cooperative performances. Where there is no visual information available from the other player, performers seem to monitor their synchronisation at a higher metrical level and thus with less attention to detailed coordination. This study, however, suggests that such synchronised behaviour may still be interpreted as being guided by some form of shared motor program (as suggested in Shaffer, 1984). However, it does not support the hypothesis that the similarity of between-hand and between-player variances and asynchronies can be accounted for by a shared motor program alone. For the highest levels of expressive and synchronised performance the performers must have some visual contact. That this is the case even when two players sit side by side suggests that in other ensemble arrangements visual information might play an even greater role. We cannot offer an explanation of the manner in which visual contact might improve cooperation, but it demonstrates its strong effect.

This study also sheds some light upon the relationship between synchronising to a fixed external time source and cooperating with another human being. When two live performers play together they seem to synchronise rather better than when they play with a nominally equivalent recording. This is somewhat surprising, since the recorded part is no more or less predictable than an ‘unseen’ partner. Such a result suggests that there is a distinction to be drawn between tracking and cooperating. In the former case the player has no ability to influence the other player, and can only predict future events, or adjust to past events. In the latter case the player may both employ this strategy but also expect the other player to adjust to their decisions regarding the timing of events. This allows the two players to exchange roles, to allow one or other to lead, whereas in a tracking condition, the live player must follow regardless.

References
Emotional expression in guitar band performance

Andrew Dry and Alf Gabrielsson
Department of Psychology, Uppsala University, Sweden

Introduction
This study is one in a series of on-going experiments at the Department of Psychology, Uppsala University, in which musicians are asked to perform short pieces of music so as to express different emotions. The general aim of this research project is to experimentally investigate which means musicians use when attempting to express various emotional characters. These means may vary depending on which type of music and which instrument(s) are used. Therefore these factors are systematically varied from one study to another. In the present study the music belongs to the rock'n'roll tradition and the instruments used are typical for this genre. For information concerning methods and results in previous investigations within the project, see Gabrielsson (1994, 1995), Gabrielsson & Lindström (1995), Gabrielsson & Juslin (1996), Juslin (1997).

To analyse music performances researchers often study various "deviations" from what is assumed to be represented in the conventional musical notation of a piece of music; for a review, see Gabrielsson (in press). This indicates that empirical research on music performance has mainly dealt with the Western classical music tradition. The methods of description and analysing art music may not be adequate when dealing with other musical forms. In order to broaden the selection of music pieces beyond the usual ones, the experiments in the ongoing research project make use of both traditional as well as modern instruments and music examples from art music, jazz, gospel, folk and pop music. The use in this paper of a rock'n'roll band's performance is in accordance with this striving. The focus will be on the use of parameters such as tempo (fast-slow), pitch level (high-low), loudness, timbre, and motion character to bring about the intended emotional characters rather than on deviations from, in this case, a non-existing notation.

The question to be investigated is: How do rock'n'roll musicians use the above-mentioned parameters - specifically the motion character and timbre - in order to express certain emotions, and how successful are they in communicating these to listeners?

Method
Four musicians were asked to perform a composed four bar verse with a pop type of melody, chords, bassline and rhythm to have it sound happy, sad, angry, solemn/ceremonial, fearful, tender and neutral. The performers were a put together Standard rock n'roll combo (vocal, guitar, bass and drums), consisting of four semi-professional young male musicians (aged 22 to 24). The experimenter acted as the bassist of the rock'n'roll combo. His specific role was to observe and participate in the performance controlling that basic musical structures were not changed, however, without unduly interfering neither with the expressive intentions of his fellow performers, nor with the process of generating the intended emotional characters. The performances took place at a local rehearsal studio and were recorded using a twelve track tape recorder with an in-built simple mixer. The recordings were mixed down to a two track audio cassette deck. The sound level of each recorded track was adjusted to avoid distortion. The recording took approximately five and a half hours. The musicians listened to each recording just after being made and were encouraged to make another recording when they were not
satisfied with the result.

In order to test that the intended emotional expressions were perceived by listeners, listening tests were conducted with 20 musically trained listeners. They rated the respective performances using adjective scales reflecting the intended emotions, furthermore adjectives reflecting various motion characters - such as striving forward, dancing, walking, rocking, swinging, uniform/smooth, rapid/fast, regular rhythmic, bouncing, floating, abrupt/stuttering, chasing and pulsating - as well as timbral properties, such as clear, full, spacious, bright, soft, dry, hard/harsh, airy, thin, close, and crispy. On the answer form each adjective was represented by a scale with numerical values from 0 (zero) to 7 (seven).

Results

The ratings obtained in the listening test were subjected to analysis of variance (ANOVA) to see whether the ratings were significantly different among the different intended expressions. Repeated measures ANOVAS in each emotion scale revealed highly significant differences among the mean ratings of differently intended expressions, $F(7, 133) = 16.30 - 33.49, p < .000000$. Tukey HSD post hoc tests showed significant differences ($p < .001$) for all pairwise comparisons within each scale, except for some cases mainly concerning differences between sad and tender performances.

Factor analysis (component analysis, varimax rotation) on the correlations between all scales was conducted and a five factor solution was adopted accounting for 60 % of the total variance. Factor I combined the tender and sad emotions with general expressiveness, with motional qualities as rocking and floating (but not stuttering), and with timbre and spatial qualities as full, soft/gentle, spacious, airy and close. Factor II combined the solemn character with walking, uniform, and regular rhythmic motion, and with clear and dry timbres. Factor III connected the angry and, to less extent, the fearful emotions with above all rapid and chasing motion (also striving forward and bouncing) and with hard/harsh timbre. Factor IV associated the happy character with dancing and swinging motion and with bright timbre. Factor V defined a thin and crispy timbre. If the factor analysis was extended to seven factors, accounting for 67 % of the total variance, factor VI essentially defined a spaciousness factor and factor VII a purely solemnity factor.

There are some general characteristics to each emotional expression that make them differ from the others. Even if a characteristic may be described with similar adjectives, the combination and the amount of interaction between them might not be the same. The main characteristics for each expression may be summarized as follows.

**Happy:** The vocalist used an irregular articulation, both staccato and legato. The performance was rhythmically soft and bouncy with emphasis on the second and fourth beat of the bar. A sort of light crisp was added with the ride cymbal. The motion character was described by the listeners as swinging, dancing, bouncing and pulsating, the timbre as bright.

**Angry:** There seemed to be an overall without-words-understanding among the musicians of how to express angry. It was performed melodically staccato and with a distorted sound. There was an emphasis on every beat, and lots of subdivisions of the beat made the piece seem go fast. Listeners rated the angry version’s motional structure as striving forward, rapid, chasing, regular rhythmic and uniform. The timbre was hard/harsh.

476
Sad: The musicians were anxious not to make it sound like a tender version. Singing was tried both in the usual octave and in an octave lower. Performers agreed that the latter was more sad, which was not what the listeners rated later. The sad version was slow and varied between the instruments. Onsets were soft and offsets long. Rhythm was rather free and glissandos were used frequently. The drum's onsets seem to be a bit delayed and with a sort of machine feel to the rhythmical execution. Listeners described the motion character as rocking and floating and the timbre as full, soft/gentle, spacious and close.

Solemn: Performers' intention was to give the impression of a pop–orchestrated national anthem with some element of march beat. The solemn version had an undynamic medium tempo with distinct changes between the tones and rhythmical emphasis on every beat. Motion character was described as walking, uniform, regular rhythmic, timbre as clear and dry.

Tender: Some of the performers' intentions were to generate delicate gentleness and closeness with connections towards tiredness and beds. The tender version was rather slow, tones were smoothed out with lyrical glissandos. It was rhythmically a bit irregular with bouncy and soft onsets and long offsets. Floating and rocking were the general motion characteristics. Timbres describing tenderness were soft/gentle, close and airy, but it was also described as spacious, crispy, full and bright.

Fearful: The performance intention was to illustrate a person being hunted and afraid to get caught. The fearful version was fast and the most dynamic both regarding tempo, phrasing and loudness. The motional qualities to describe fearfulness were striving forward, rapid, chasing, stuttering, irregular rhythmic and not uniform/smooth. Timbre was thin, crispy, and hard/harsh.

Neutral: There were no specific comments on how the musicians would perform this version. They simply played it as if it were a song that they just learned. The neutral version was in medium tempo, rhythmically regular and staccato with short on- and offsets. Its motion was described as uniform/smooth and regular rhythmic and the timbre as clear and bright, dry and airy.

When comparing these results with other investigations in this series (such as Gabrielsson, 1994; 1995; Gabrielsson & Juslin, 1996; Gabrielsson & Lindström, 1995; Juslin 1993, 1997; Lindström, 1992) there are some striking similarities but also some differences. The musicians' expressive intentions seemed to be mostly intuitive and non-verbal, just as with the electric guitarists in Juslin (1993) "were not consciously thinking about what means they were going to use in order to produce each emotional character". Regarding the mean tempo results just as in the other investigations angry and fearful versions were the fastest, sad and tender versions the slowest. The happy version in this investigation was not so fast as has been the case in the earlier investigations.

In the earlier studies there has been a problem to distinguish between sad and tender versions. The musicians were aware of this problem, and the listeners' ratings showed differences between these two expressions. Still, however, the ratings of these two emotions were rather close, and in the factor analysis both ended up in the same factor. On a general level sad and tender expressions seem to have similar motional and timbral features, but the amount of each ingredient is not the same.

An interesting result was the connection between angry and fearful, for instance, appearing together in Factor III of the factor analysis and associated with qualities as striving forward,
rapid, chasing, bouncing, and hard/harsh. Earlier results showed that performances of anger and fearful include non-legato articulation and irregularities in tempo (Gabrielsson and Juslin, 1996). The musicians' expressive intention concerning fearful was to perform the music as a scene where someone was hunted. The listeners adequately understood that it was the hunted person's perspective rather than the hunters' that was the intended expression, even if the music in itself could in some way illustrate the aggressive hunters.

Discussion
One might question the experimenter's role in connection with the performances. Acting as one for the musicians, he might have influenced the performance by different unintended paralingual or kinaesthetic cues, such as variations in voice or facial expressions, even if he did have careful thought of how this could be avoided. Of course, a recording situation according to the instructions used is rather unusual and may generally have affected the performers' behaviour in different ways. Do rock'n'roll musicians even think in terms of communicating different emotions? Or is it that they compose a song that reveals some feelings that they later observe as listeners? On the other hand these results do not indicate anything that could be considered as news for any musician. Even if these performances would be musical clichés, they certainly managed to make the listeners understand what they were trying to express—which is just what the musicians were asked to do.

Acknowledgement
This research was supported by the Bank of Sweden Tercentenary Foundation.

References
Emotional expression in singing: A case study

Girilai Baars and Alf Gabrielsson
Department of Psychology, Uppsala University, Sweden

Introduction
The performance of music has been studied from many different points of view, from post-hoc discussions of the performances of whole scores to experimental analyses of single notes or intervals and simple rhythmic patterns. Almost all the research uses tonal Western art music as a reference point, and little has been done on vocal performance. An early pioneering study of singing is that by H. G. Seashore (1937) in which he analyzed, in detail, nine singers' performance of arias from Handel's *Messiah* and other pieces.

This study is part of a series of studies aimed at investigating the means that music performers use to express and communicate emotions. Professional musicians are asked to perform short pieces of music in order to express different emotions: happiness, sadness, anger, fear, tenderness, solemnity and a version "without expression." Various instruments have been used: violin, flute, saxophone, electric guitar, synthesizer, sentograph, and the singing voice (e.g., Gabrielsson, 1995; Gabrielsson & Lindström, 1995; Gabrielsson & Juslin, 1996; Juslin, 1997). Listening tests are conducted to see whether the intended expression are perceived by listeners. These studies have demonstrated that practically all performance variables - such as tempo, timing, articulation, loudness, amplitude envelopes, timbre - are affected by the different intentions concerning emotional expression. The means that are used depend, of course, on which instrument is used. The singing voice is an extremely flexible instrument capable of manipulating all the parameters mentioned above. Its main restriction is the range of pitches within a singer's reach.

There are as yet few controlled studies of emotional expression in singing. Kotlyar and Morozov (1976) had eleven singers perform phrases from songs or arias and express happiness, sorrow, fear, anger, and a neutral version. Most of these were correctly identified by judges. Tempo was fastest for fear, slowest for sorrow. Anger was performed loudest, fear softest. Tone onset was slow in the sorrow version and fast in anger and fear versions. Baroni and Finarelli (1994) had three actors and three singers utter and sing, respectively, short phrases from opera arias. The phrases should be perceived as serene and joyful, sad and depressed, and aggressive and were associated with certain scenarios. Some differences between the performances by actors and singers were described and discussed. Ohgushi and Hattori (1996) found that emotional communication in singing was not only enhanced by visual information from the singer (facial expressions, etc), but that pure visual information was easier to interpret than pure auditory information. Recently, Rapoport (1996) reported a detailed study of emotional expression in opera and lied singing. He identified over 50 types of vocal tone structures which were classified into eight categories.

Singing performance has only been studied slightly in the on-going research project (Gabrielsson & Juslin, 1996). In the present study a female singer, mainly active in the folk music genre, was asked to perform and express various emotions according to procedures similar to our earlier investigations with instruments. By employing a folk singer, a fresh area of musical activity is scrutinized and the data gathered will help widen the scope of this line of research.
Method
A professional folk music singer was asked to perform two songs: *What Shall We Do With the Drunken Sailor* (with lyrics) and *Greensleeves* (without lyrics, free syllables) using a series of different expressions: Happy, Sad, Angry, Tender, Fearful, Solemn, and No Expression. The performances were recorded on a tape recorder and then sampled at the rate of 22 kHz into a computer. A listening test involving nine musically qualified judges was performed in order to see whether their perceptions matched the singer's intentions. The listeners graded each performance on a 0–10 scale for seven adjectives matching the seven intended expressions. The performances were analysed in order to study how parameters such as tempo, intensity, articulation, duration of measures, etc., were used in the singer's different performances. This analysis was made using the Sound Swell program (Ternström, 1992). The resulting data were analysed with the Rhythm Analyser program (Madison, 1992).

Results

Listening Tests
Statistical analysis of the listeners' ratings (repeated measures analyses of variance) showed that their ratings differentiated between the different versions in accordance with the performer's intentions. However, there was some confusion between the sad and tender versions. The sad version of *Greensleeves* was rated as being very sad and tender, and the tender version was likewise rated very sad and tender. The tender version of *Drunken Sailor* also received a high sad rating. On the other hand the tender and sad adjectives in the two respective versions were rated significantly different from the other adjectives.

Performances

Tempo. Tempo of the different performances are shown in Table 1. The fastest performances for both songs were the angry versions; they were followed closely by the happy and fearful versions. The sad versions were the slowest.

<table>
<thead>
<tr>
<th>Table 1: Ranking of tempo (bpm) for Greensleeves and Drunken Sailor</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GREENSLEEVE</strong></td>
</tr>
<tr>
<td>Angry (70)</td>
</tr>
<tr>
<td>Happy (52)</td>
</tr>
<tr>
<td>Fearful (48)</td>
</tr>
<tr>
<td>NoExpression (44)</td>
</tr>
<tr>
<td>Tender (35)</td>
</tr>
<tr>
<td>Solemn (32)</td>
</tr>
<tr>
<td>Sad (28)</td>
</tr>
</tbody>
</table>

Articulation. The most legato version of *Greensleeves* was the no expression version, followed by the sad and tender versions. There were no staccato performances, but fearful is the least legato, followed by the angry version. However, the most legato versions of *Drunken Sailor* were angry and happy, which is contrary to the results from earlier studies. Least legato were fearful and sad, the latter being unexpected, too.

Sound Level. The angry versions were predictably the performances with the highest sound level, and the fearful were the ones with the least sound level. In fact, the sound level rates for the two songs matched each other rather well, with the no expression versions found at the
center of the rankings. Some of these results may be studied in Figure 1, which shows the amplitude envelopes for two versions of each song. Figure 1 (top left) demonstrates the high sound level of the angry version of *Greensleeves*, but also the variation in dynamics, whereas the no expression version in Figure 1 (top right) looks almost uniform in comparison. The happy version of *Drunken Sailor* (Figure 1, bottom left) looks similar to the angry version above, although at a lower sound level. The fearful version was by far the softest of all, as convincingly demonstrated in Figure 1 (bottom right).

*Figures 1: Amplitude Envelopes for Greensleeves: Angry (top left), No Expression (top right) and Drunken Sailor: Happy (bottom left), Fearful (bottom right).*

**Timing.** Only a few aspects of timing have been selected for analysis, mainly due to the vast amount of data that a complete and detailed approach would generate. The performances were compared to each other with regard to deviation from the nominal values given in the notation, on a measure–by–measure level. The ratios of simple recurring rhythmic patterns (e.g. dotted eighth note : sixteenth note) were also compared.

The fearful version of *Greensleeves* showed a great deal of variation at the measure level, as did the tender and sad versions. *Drunken Sailor* showed a slightly different pattern, with the sad and tender version as the most varied, and the fearful version as one of the more regular performances. The no expression performances were the most regular for both songs.
The mean ratios of the recurring quarter note : eighth note patterns in Greensleeves were below the nominal 2:1 ratio in all versions, the means ranging from 1.52:1 for the angry version and 1.57:1 for the fearful version to 1.95:1 for the sad version. The version with no expression had a mean ratio of 1.75:1. On the other hand the ratio of the dotted eighth note : sixteenth note pattern in the second half of most measures of Greensleeves varied between 4.34:1 for angry and 2.81:1 for happy. Only the happy and fearful (2.94:1) versions came close to the the nominal 3:1 ratio, and all the other versions had higher ratios (3.21:1 to 4.34:1)

Embellishments. The performer also used some genre specific tools of expression, such as the way she employed embellishments and ornaments. The angry, fearful and no expression versions of both songs had no embellishments, or very, very little. Drunken Sailor generally had less embellishments, presumably because the use of words permits less variation/deviation than the free–syllabic singing in Greensleeves. When embellishment was used in Drunken Sailor, it was at the very end of the song, always in "early in the morning," particularly in "early."

In Greensleeves the performer used two different kinds of embellishments. In the solemn and sad versions accented, long notes often started one semitone low, followed by a leap up to the correct pitch. In the happy and tender versions accented, long notes often started on the correct pitch, to be followed by a leap or sometimes a series of leaps up, and then back to the original pitch. The interval was often a semitone, but sometimes more.

The performer did not use vibrato as an expressive tool at all. But there was a great variation of timbre, ranging from the childish glee in her voice when singing the happy version of Drunken Sailor, to the droning, robot-like voice of the no expression version of that song.

Discussion

By and large, these results confirm earlier results obtained in the project (e.g., Gabrielsson, 1995; Gabrielsson & Lindström, 1995; Gabrielsson & Juslin, 1996). The results regarding tempo and sound level match previous findings. Regarding articulation there are some discrepancies in comparison to earlier results. These could well be due to the fact that the use of articulation in singing performance is probably different from its use in instrumental performance.

The regularity of the performance of the fearful version of Drunken Sailor differed markedly from the highly irregular performances of fearful that have been registered in our previous studies. However, the fearful version of Greensleeves was in fact highly irregular, thus confirming earlier findings.

The two different kinds of embellishments used in Greensleeves, do show a certain consistency: the leap up to pitch from a semitone below – in the sad and solemn versions – has a more formal sound than the more improvisational embellishments used in the tender and happy versions. And perhaps the solemn and sad emotions are a touch more formal than happiness and tenderness.

The lack of vibrato is a typical feature of Swedish folk music, as is the use of embellishment. Furthermore, the rather wide ranging differences in timbre that the performer employed can also be explained by the genre of the performer – the classical tradition of singing does not normally allow for that kind of expression, relying on other expressive tools instead.
Acknowledgements
This research was supported by The Bank of Sweden Tercentenary Foundation.

References


1. Introduction

Although the relationship between music and emotions has been widely studied since last century, a relatively few studies tried to link the interpretative aspects of music performance with the expression of single emotions. Music performance has mainly been studied within the context of musical structure: different levels of interpretations have been described according to the skills available to the performer (Sundberg, 1991 for an overview), and different interpretative rules have been indicated in systematic "deviations" from the musical score (Sundberg, Askenfeld, Fryden 1983). In this search, some important aspects of emotional experiences, such as emotional control and regulation, have not always been taken into consideration. In some recent research on this field, it was possible to link musical structures to peak emotional experiences (Sloboda, 1991).

The importance of interpretative elements as means to communicate and express emotions beyond the musical written text has been investigated by Gabrielsson (1994, 1995) and quite recently by Gabrielsson and Juslin (1996). In this last paper violin, flute, singing and electric guitar interpretations were compared. An encoding-decoding procedure was adopted: players or singers were invited to interpret a piece according to a specific emotion and then the same interpretation was presented to a group of judges in order to assess if it was correctly recognized. This procedure, which is quite usual in the classical studies concerning facial and vocal expression of emotions, is rather new in the music field. It has been adopted in Italy too in a study by Canazza, De Poli and Vidolin A. (1996) on clarinet expressivity and it seems to have produced promising results, because a double analysis of musical or sound features and emotional components can be made more precisely than in a free listing situation.

In the study of Gabrielsson and Juslin (1996), singing resulted to have more limits than instrumental playing in the expression of single emotions. Different technical possibilities and a different control of vibrato parameters (Sundberg, 1987) can be indicated as some reasons, but we must not omit the fact that voice belongs to our body and has many functions which go beyond singing. In this respect a comparison can be made with the studies on spoken voice and its emotional quality (Scherer, 1986). Scherer's model can be interesting in many ways. Vocal expression of emotions is considered in its regulatory aspects: single emotions are built around different parameters in so far as they express different regulatory attitudes.

The quality of voice is somehow influenced by coping strategies, self evaluation and evaluation of environment. A voice can be wide or narrow, relaxed or tense, full or thin in relation with the possibility a person has of fulfilling his or her goals and wishes. Alongside Scherer's works, we may consider other stimulating indications in a work of Anolli and Ciceri (1992), which takes into account the possibility of applying methods and results on spoken voice also in the field of the singing voice. Also Scherer (1991) has recently discussed the latter point. Sundberg, Iwarsson and Hagégård (1994) made a comparison between two different performances of sung phrases: one "appropriate" and another void of emotional expression. Scherer and Siegwart (1993) tried to compare different interpretations of singers in order to map single emotional elements and their relation to musical and acoustic parameters. When we consider singing, we could either assume that some regulatory functions of spoken voice are present or assume that a different regulatory activity is going on, more close to the re-evocation of emotion rather than to its direct expression. In any case, we have to stress that voice, body and emotion are closely connected. Therefore, when we sing we encode our emotions not only in our voice, but also in our body, in our face and, above all, in our brain.
2. Method

**Encoding and Decoding Procedures.** Three professional singers and three professional actors were asked to interpret three different phrases according to three different emotions: happiness, anger, sadness. The phrases, taken from famous operas, were chosen because of their semantical aspect that, although not "neutral", nevertheless permitted different kinds of emotional intonations: 1) «Dunque al giardin verrai?» [So will you come to the garden?] (Mozart, *Le nozze di Figaro*, act 3, scene 2); 2) «Ed or fra noi parliam da buoni amici» [And now let us speak as good friends] (Puccini, *Tosca*, act 2, scene 5); 3) «Nulla sfugge al mio sguardo» [Nothing escapes my view] (Verdi, *Aida*, act 3, scene 3). Scenarios for each emotion were imagined in order to give to each of them well-defined characters. The singers' performances were provided with piano accompaniment which was not taped (the singers could listen to the music only through their headphones).

A first version of the present research was presented at the ESCOM Conference of Liège (Baronin-Finarelli 1994). In a subsequent phase of the research, three independent judges evaluated the taped performances on two scales: a five point scale to assess the recognizability of the expressed emotions (1=very poor; 2=poor; 3=fair; 4=clear; 5=very clear) and a four point scale to assess how genuine and natural were the single interpretations (1=quite exaggerated; 2=rather forced; 3=almost genuine; 4=genuine). As a result of this evaluation the two best interpretations of singers and actors were chosen. Two tapes (A and B), containing the chosen performances of actors and singers respectively, were submitted to 42 University students. The students had to mark on an answer sheet the name of the emotion (happiness, sadness, anger) attributed to the recorded stimulus.

**Main Hypothesis.** These are our main predictions: 1) All the stimuli of tapes A and B should be recognized in percentages above the random level (33.33%); 2) stimuli of tape A should be recognized better than stimuli of tape B as actors have more freedom in their interpretation in comparison with singers who have to follow the prescriptions of the written music; 3) differences among emotions can be predicted according to the results in literature: in the vocal channel, happiness and positive emotions are not recognized so well as anger and sadness; 4) it could be possible to describe specific emotional profiles both in spoken and singing phrases according to acoustic parameters, sound quality and duration.

**Data analysis.** We measured fundamental frequency and micro- and macro- variations of loudness, pitch and durations of phrases, syllables, notes, in order to define the parameters involved in the expression of single emotions. Timbral analysis and modifications of formants of the singing voice are still under discussion and will not be presented here. Audio files were sampled at 44.100 kHz (resolution of 16 bits) with hard-disk recording system Digidesign Sound Tools II. Analyses were made with Sound Scope (GW Instruments) on Apple Macintosh PowerPC.

3. Results

The results go in the direction of our predictions: all but one of our stimuli were correctly recognized with percentages far above the random level. Emotions expressed by actors are more easily recognized than emotions expressed by singers (Table 1). All the differences are highly significant (p<.001). There is a significant difference between the recognition of happiness and other emotions, either when expressed by actors or by singers (p<.001). But if we consider (Table 2) the different phrases, we can see how the “happy” interpretation of “Nulla sfugge al mio sguardo” in the singers was very poor. The different results in recognition of spoken and sung phrases, and particularly spectacular differences such as that referred to “Nulla sfugge”, attracted our attention and needed an explanation. Our explanation was that the reasons for the difference were not to be referred only to the limits of freedom of the singers, but to the composers themselves. In fact, they had already given to the phrase a precise emotional connotation. And this fact deeply altered the interpretation of spoken and sung performances.
Table 1. Recognition of emotions in spoken and sung voice. Percentage Data.

<table>
<thead>
<tr>
<th>Emotions</th>
<th>Actors</th>
<th>Singers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td>81.83</td>
<td>51.16</td>
</tr>
<tr>
<td>Anger</td>
<td>96.00</td>
<td>75.16</td>
</tr>
<tr>
<td>Sadness</td>
<td>94.50</td>
<td>77.66</td>
</tr>
<tr>
<td>All the emotions</td>
<td>90.72</td>
<td>68.00</td>
</tr>
</tbody>
</table>

Table 2. Recognition of actors (A) and singers (B) interpretations of single emotions and phrases. Percentage Data.

<table>
<thead>
<tr>
<th>Phrases</th>
<th>Happiness</th>
<th>Anger</th>
<th>Sadness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dunque al giardin...</td>
<td>86.35 (A)</td>
<td>91.15  (A)</td>
<td>86.35 (A)</td>
</tr>
<tr>
<td>Ed or fra noi...</td>
<td>66.65 (A)</td>
<td>95.80  (A)</td>
<td>98.60  (A)</td>
</tr>
<tr>
<td>Nulla sfugge...</td>
<td>93.05 (A)</td>
<td>100.0  (A)</td>
<td>98.60  (A)</td>
</tr>
</tbody>
</table>

So our singers were not asked (as were the actors) to give an emotional expression to a neutral phrase, but to redefine and even to contradict the emotional character already present in the musical phrase and given by the composer. It was just this contradiction or in any case the interference of different emotional meanings that made the recognition more difficult. Before interpreting the performance, it was necessary to interpret the musical phrase itself.

The emotional or semantic interpretation of a written musical text is a considerable and yet unsolved musicological problem: interpretation is always ambiguous and probably the nature itself of music (and of every kind of art) need ambiguities and cannot avoid them. Nevertheless some orientations in hermeneutic activities do exist: studies in this field have been developed by semiotics and by psychology itself (Sloboda 1996). In our case we looked for some well known schemes which could provide arguments supporting our intuitive hypotheses.

Dunque al giardin verrai:
- rhythm: meter Allegro 4/4, with regular and rapid notes (all of 1/8);
- contour: ascending by steps (descending skips are only a feminine ending);
- tonality: maximum stability (beginning and ending on A with harmony I-V-I).

According to Imberty (1988) a musical model like this reflects a “formal integration” corresponding to “psychic integration”, with a relevant rhythmic dynamism. Mozart interprets the phrase as a confirmation of joyful expectations.

Ed or fra noi parliam
- rhythm: Andante moderato 6/8, with regular pattern of 2/8 + 1/8 and final prolongations;
- contour: regularly ascending /descending;
- tonality: slight instability (beginning and ending on V degree) with slight harmonic mobility reducible to I-V-I.

Again “formal integration”. But ondulatory contour and 6/8 rhythm allude for intertextual reasons (Agawu 1991) to the romantic intimistic model of a barcarole. In the score Puccini adds: «con galanteria».

Nulla sfugge al mio sguardo
- rhythm: Allegro vivo 4/4 with differentiated durations and short final notes;
- contour: one repeated note with an ascending inflexion on the last accent;
- tonality: instability and tonal ambiguity; dissonant harmony.

Clear symptoms of what Imberty calls “formal disintegration”; disintegration can have two aspects: aggressive-anguished or depressive-melancholic. For its rhythmic vitality, the phrase can be assigned to the former aspect. One can add also that in Verdi’s tradition (Dalmonte 1979) recitative on an only note is always linked to a situation of threat.
Singers manifested evident difficulties in modifying pre-defined emotional expressions. Now it is clearer why no one of them succeeded in singing *Nulla sfugge* "happily". But we can notice also that the sentimental gallantry of *Ed or fra noi parliam* cannot be easily given a frankly aggressive or a sad character: anger was transformed in a sort of mockery and sadness in a sort of imploaring request. Recognition was possible because it was relatively easy to collocate the examples into one of the three proposed categories, but their acoustic analysis showed that their characters were different from those of the corresponding examples of the same category.

In Figs. 1 and 2 we give the results of our analysis for five phrases chosen from the better recognized nine ones (see Table 2). We do not give the results for *Ed or fra noi* nor for the "happy" *Nulla sfugge*. Fundamental frequency was obtained through a procedure of pitch-tracking that allows to follow in real time a chosen amplitude peak (in our case the fundamental one) within a region of the frequency spectrum. This procedure permits also a good observation of vibrato amplitude. In the graphic examples, interrupted lines are traced due to unvoiced zones or to the presence of unpitched consonants (r, s, t).

In the comparative analysis of the Figures we took into consideration the following parameters:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Envelopes of each syllable</th>
<th>Global dynamic sequence</th>
<th>Deviations from correct pitch</th>
<th>Glissandos and vibratos</th>
<th>Regularity/irregularity of rhythm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loudness</td>
<td>1. envelope profiles</td>
<td>2. global dynamic</td>
<td>1. deviations from correct</td>
<td>2. glissandos and</td>
<td>2. regularity/irregularity</td>
</tr>
<tr>
<td>Rhythm</td>
<td>1. global duration of the</td>
<td>sequence</td>
<td>pitch</td>
<td>vibratos</td>
<td></td>
</tr>
<tr>
<td></td>
<td>phrase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Regularity/irregularity was not deduced directly from the graph but from the measure of all the micro-deviations from the prescribed durations. The global rate of deviation was computed independently from the shortening or prolongation of the global durations.

The following Table 3 gives a shortened summary of the principal results of our analysis:

<table>
<thead>
<tr>
<th></th>
<th>Rhythm</th>
<th>Loudness</th>
<th>Pitch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>Shortened global duration</td>
<td>Shortened attacks</td>
<td>Very imprecise pitch</td>
</tr>
<tr>
<td></td>
<td>Regular single durations</td>
<td>All notes accented</td>
<td>Reduced gliss. vibr.</td>
</tr>
<tr>
<td>Sadness</td>
<td>Prolonged global duration</td>
<td>Gradual attacks</td>
<td>Low pitch tendency</td>
</tr>
<tr>
<td></td>
<td>Irregular single durations</td>
<td>Irregular accents</td>
<td>Frequent gliss. vibr.</td>
</tr>
<tr>
<td>Happiness</td>
<td>Regular durations</td>
<td>Gradual attacks</td>
<td>Correct pitch</td>
</tr>
<tr>
<td>(single and g local)</td>
<td>(single and global)</td>
<td>Regular accents</td>
<td>Some gliss. vibr.</td>
</tr>
</tbody>
</table>

4. A final remark

In Table 3, we did not mention the relationships between spoken and sung versions of the phrases. In this paper we cannot adequately discuss this important theme. We limit ourselves to a shortened remark about some of the differences between the two kinds of expression. First of all, it is difficult to compare directly our results with those of speech studies, for example with Scherer's findings. In singing, for example, particular parameters (such as vibrato or glissando) have a great importance but are not so important in spoken voice, while for some other parameters (such as fundamental frequency) the situation is quite the opposite. Nevertheless, some aspects of sung phrases can be traced back to dimensions described by Scherer (1986): for example his "hedonic" quality could be linked (in sung expressions) to regular or irregular patterns in the rhythmic, dynamic and intonational elements. A future perspective of our research on sung performance will take into consideration problems not mentioned in the present occasion: above all the relation between sung and spoken utterances and the analysis of the timbral quality of voice. Another crucial point ought to be the analysis of musical and vocal aspects in relation to facial, gestural, physiological expressions.
Dunque al giardin verrai - Pitch and amplitude envelope

HAPPINESS

total length
sec. 2.303

ANGER

total length
sec. 2.134

SADNESS

total length
sec. 2.886
Nulla sfugge al mio sguardo - pitch and amplitude envelope

References


Dalmonte R. (1979), Da "Oberto" a "Rigoletto": precisazione di una formula, "Ricerche musicali", 3.


Emotion and Expression in Opera and Lied Singing
- From Sound Analysis to Brain Processes

Eliezer Rapoport
Musicology Department, Bar - Ilan University, Ramat-Gan, ISRAEL

March 16, 1997

Abstract

While music perception and cognition are concerned with acoustical signal processing in the listeners' brain and mind, the present author's work is concerned with vocal music performance, centered on the performing artist and his/her ways of conveying emotional messages to the audience. To this end FFT spectrograms of the vocal tones are studied and the various kinds of information encoded in them are analysed and deciphered, to the utmost detail. This is a scientific-objective tool, independent of personal biases. From this work, a very systematic classification of vocal tones, and the various ways emotional characteristics are encoded into them in singing has emerged, which was then related to basic neurophysiological processes in the singer's brain. These processes operate in translating the emotional messages of the text and the musical score into musical tones as uniquely shaped acoustic signals by characteristic commands sent from the brain as neuron firing sequences activating the musculature that controls the vocal folds tightening and vibrating frequency.

In recent papers [Journal of New Music Research 25, pp.109-149, (1996), Proc. XI Colloquio di Informatica Musicale, Bologna, 1995], the author demonstrated that emotional messages in singing are encoded at the level of the single individual tone as characteristic temporal structures constructed from various characteristic expressive elements. These are actually timbre elements, as they use specific parts of, or add extra frequencies to the ensemble of frequencies contained in the vocal tone. The Fast Fourier Transform (FFT) analysis was shown to be very potent in the identification and deciphering of the emotional expression code in vocal tones. These expressive timbre elements are: (1) singing in the frequency range corresponding to the lower formants - phonation in the lower formants, (2) excitation of the higher partials of the singing formant, (3) vibrato, (4) transition - a gradual pitch increase from the onset to the sustained state, (5) pitch change within the tone, (6) sforzando - an abrupt pitch increase at the onset of the tone, and (7) unit pulse.

Excitement or calmness were shown to be built into the vocal tones in such a way that the smaller the number, and the more gradually these aforementioned timbre or expressive elements enter into operation - the more calm the tone will be, and.

490
inversely: the more abrupt, and the larger the number of the timbre elements operating simultaneously - the more excitement is built into the vocal tone.

On this basis, the author was able to classify the large number and variety of (temporal) structures or (acoustical) shapes of vocal tones encountered in singing of western art music (opera arias and lieder) into eight categories or tone families in a very well-defined hierarchical scheme.

The fundamental frequency (or pitch) of the vocal tone is determined by the length and mass, but predominantly by the momentary tension in the singer's vocal folds, in such a way that increasing tension leads to increasing pitch. In singing the singer constantly changes the vocal folds tension in order to produce the various tones. The Fast Fourier Transform (FFT) spectrogram, displaying frequencies - time diagram is thus equivalent to the vocal folds tension - time diagram. The author identified a basic unit, appearing as the inverted letter U in FFT spectrograms, with time duration of 100-160ms. This unit, to be called "unit pulse", is actually a frequency increase followed by frequency decrease, corresponding to vocal folds tension increase followed by tension release. This can be traced back to some basic unit command in singing sent from the brain to the vocal folds (via the central nervous system and the musculature activating the vocal folds). Each vocal tone is composed of one or more unit pulses, and always an integral number of pulses. The unit pulse is therefore a unit of brain time counting (pacemaker or biological clock) operative in rhythm.

Excitement and calmness in singing can be defined on a neurophysiological level of tightening of the vocal folds: strong and rapid tightening command corresponds to excitement, whereas weak and slow tightening corresponds to calmness or relaxation. This leads to redefinition of excitement and calmness on the 100-160ms timescale in terms of three varieties: large pulse, small pulse, and "zero pulse", their various ways of their arrangement within the vocal tone (temporal) structure reflect the singer's ways of expressing emotions in singing. The author proposed a simple neuro-physiological unit pulse model for singing, and the unit pulse was demonstrated in a number of opera arias and lieder.

The well-documented observation by many authors [the groups of Sundberg in Sweden, and of Castelengo in Paris] that in legato coloratura singing the behavior of each individual tone consists of the unique and characteristic feature of initial pitch increase, with an overshoot above and beyond the nominal intended tonal pitch, followed by immediate pitch decrease below the intended nominal pitch, with the pitch perceived consisting of a time integration over this peculiar pitch-time trajectory, is very consistent with the unit pulse model. Thus legato coloratura singing consists of a sequence of unit pulses.

Vibrato is also interpreted as a pulsetrain of unit pulses. Some measurements on the behavior of individual vibrato pulses along a single tone and along a melodic phrase - are to be presented and discussed here.

Indeed, recent physiological electromyographic (EMG) measurements[Titze and coworkers] demonstrated the correlation of vocal vibrato with characteristic pulse sequences of motor unit firing of well defined frequency modulation patterns in the cricothyroid muscles. This muscle activity is controlled by corresponding brain com-
mands of neuron firings of well defined frequency modulation patterns. More recent neurological measurements on bird singing elucidate in more detail the brain mechanisms involved in bird singing [Yu and Margoliash].

The concept of "fastest pulse" or "elementary pulse", as deduced from ethnomusicalogical studies of African polyrhythmic music, was discussed in a recent review article on rhythm perception [Seifet et al.]. Ideas concerning the possibility of existence of some unit of motor action in the field of human motor behavior also appear in the literature [Viviani]. Thus, the unit pulse in singing might belong within a family of more general processes.

In the present study unit pulse parameters (in the frequency-time diagram), such as pulse height, duration, rise time, etc., were measured, with the goal of seeking correlation with the level of excitement in singing. Some combination of these parameters might eventually yield a single parameter to serve as a "pulse strength" and be a quantitative measure of the degree of excitement. In the present report, the pulse height, and a parameter related to the rate of tightening of the vocal folds, were examined.

A very simplified description of singing was found, in terms of three entities: strong pulse, weak pulse, and "zero pulse", and a vocal tone beyond a certain excitement threshold is described as a sequence of pulses with the arrangement of either small pulses followed by increasingly larger pulses - this gradual structure is typical of calm and expressive singing, whereas a sequence of a very large first pulse followed by other large pulses, or by pulses of steadily diminishing strength is characteristic of excited singing. This is very similar to computer "machine language" in the binary system of zero and one.

Analysis of performance of vocal music in terms of the unit pulse allows quantitative measurement of the momentary degree of excitement for each individual tone along a sung melodic phrase, and gives an insight into the singer’s intentions and personal ways of expression and interpretation.
How Can Music Performers Become Better at Communicating Emotions to Listeners?

Patrik N. Juslin

Department of Psychology, Uppsala University, Sweden

This paper constitutes part of an attempt to transcend the traditional score-based accounts of music performance, in which performance expression is seen as resulting from the performer's structural interpretation. According to the functionalist perspective (e.g., Juslin, 1995; 1997a), humans (a) have an innate capacity for the nonverbal communication of basic emotions, and (b) may exploit this capacity in music performance. A number of recent studies have shown that professional performers are able to communicate iconic representations of specific emotions (e.g., "anger", "sadness", "happiness", and "fear") through their performances of a piece of music (cf. Gabrielsson, 1995; Gabrielsson & Juslin, 1996; Juslin, 1997a). Performance analyses have revealed that performers use a number of expressive cues in their performances, such as tempo, loudness, articulation and timbre, in order to generate the desired emotional expression. Similarly, experiments using synthesized and systematically varied performances have indicated that listeners use these same cues in judgments of the emotional expression (Juslin, 1996b). Both encoding (expression) and decoding (recognition) aspects have thus been studied, but so far no systematic attempt has been made to relate the two aspects. However, without a joint description of the cue utilization of performers and listeners, it is difficult to understand why the communicative process has been successful or not in a particular situation. This knowledge is, of course, essential if one wants to improve the communicative process. Sloboda (1995) has recently stated the need for applicable music psychology in a provocative way:

"Suppose that at this instant every psychologist of music, and every piece of writing that any of them had published, were to instantly vanish and be lost from memory. How, if at all, would this diminish the quality and the quantity of music experience in the world...Would it lead to performers or composers to carrying out their tasks less well?" (p. 20)

How, then, may the knowledge from our studies of emotional expression in music performance be used by practicing performers or listeners? The purpose of this paper is to go beyond a mere description of the communicative process, and to point to ways of actually improving it. This leads to the necessity for a methodology which can (a) provide a simultaneous understanding of the performer's and the listener's cue utilization; (b) render that understanding open to public inspection; and (c) provide procedures for the improvement of the communicative process. The functionalist approach (Juslin, 1995; 1997a) is uniquely suited to this purpose. It involves the integration of ideas from research on emotion and nonverbal communication with Brunswik's (1956) metatheory. Part of this framework is a modified version of his lens model intended for communication in music performance. The lens model illustrates how the performer encodes an emotion by means of a number of expressive cues in the performance (e.g., tempo, loudness, and articulation). The resulting emotional expression is decoded by the listener, who uses these same cues to judge the intended expression. However, none of the cues are fully dependable as predictors of the intended expression. As evidenced by empirical findings, the cues are only probabilistically related to the performer's expressive intention (Juslin, 1997a). This is because (a) there are wide inter-individual differences in cue utilization among performers, (b) different cues are available on different instruments, (c) cue utilization is constrained by the structure of particular melodies, and (d) cues may be used in the same way in more than one expression. In other words, there is large variability in code usage across performers, listeners, instruments, and melodies. Nevertheless, performers are rather successful in their attempts to communicate emotions to listeners. This seemingly contradictory state of affairs is actually predicted by the
functionalist perspective: High accuracy in spite of large variability in code usage is explained by reference to the Brunswikian (1956) concept of *vicarious functioning*, i.e., the fact that the process involves several probabilistic but partly redundant cues that listeners have to combine in flexible ways in order to arrive at reliable judgments of the emotional expression. In the lens model, the statistical relationship between the performer's expressive intention and a particular cue in the performance describes the *ecological validity* of that cue. The *functional validity* of the cue is indexed by the statistical relationship between the cue and the listener's judgment. The lens model has some important implications. First, since the cues are only probabilistically related to the performer's expressive intention, it means that accuracy cannot be more than probabilistic. That is, we cannot expect perfect communication. Second, because the cues are intercorrelated, many different cue utilization strategies may lead to a similar level of accuracy. This is because there is no pressure towards uniformity (cf. Brehmer, 1994).

How can we describe the complex relationships among performers, cues, and listeners? Ideally, we would like to capture both the performer and the listener systems using parallel concepts (Brunswik, 1956). This may be accomplished through the use of multiple regression analysis (MRA). MRA is a data-analytic system sufficiently flexible to handle the complexity (multiple cues, uncertainty, intercorrelations among cues) of the communicative process. It permits the analysis of the joint and the unique influences of multiple independent variables on a dependent variable, and provides *effect size* measures that are easily understood and communicated (Cohen & Cohen, 1983). Therefore, multiple regression models may provide us with intuitively understandable answers to the following questions (which may be asked for performers and listeners alike): (a) Which cues are used? (b) What is the relative importance of the different cues? (c) Is a cue used linearly or nonlinearly? (d) How is the information from different cues combined? (cf. Brehmer, 1994). Thus, it seems suitable to apply regression models to both the performer and the listener systems. But how, exactly, should the two systems be related? The solution to this problem was the introduction of the Lens Model Equation (LME) by Hursch, Hammond, and Hursch (1964). In the following, the LME is described with respect to how it applies to communication in music performance. For information on how it was originally conceived in studies of human judgment, the reader is referred to Cooksey (1996). The Lens Model Equation (eq.1) embodies the truism that *achievement* \( r_a \) (that is, the correlation between the performer's intention and the listener's judgment) is a function of *cognitive control* \( r_c \), *ecological predictability* \( R_e \), and *matching* \( G \). Ecological predictability and cognitive control refer to the multiple correlations of the regression models of the performers and the listeners, respectively. These indices reflect the extent to which the linear models fit the data, and may be interpreted as measures of the consistency of cue utilization. \( G \) is a measure of the extent to which the cue weights of the performers and the listeners are matched to each other.

\[
(\text{eq.1}) \quad r_a = G R_e R_s 
\]

For practical reasons, most performance studies include only a limited number of melodies and performers (Palmer, 1997). This makes it difficult to apply multiple regression analysis, where one as a rule needs ten times as many cases (i.e., performances) as the number of cues. If fewer cases are included, the cue weights will most likely be unstable and unlikely to replicate. In this study, the problem was solved by integrating a large body of (published and unpublished) data from several studies into a larger analysis comprising 46 performances. The purpose was to study (a) the achievement of the communicative process, (b) the ecological predictability of the performers, (c) the cognitive control of the listeners, (d) the cue weights of the performers.

---

1However, there are empirical findings that suggest that the usual statistical procedures for estimating the stability of regression weights may be too conservative for Judgment Analysis (Cooksey, 1996).
and the listeners, (e) the extent to which the performers' and the listeners' relative weights were matched, and (f) the intercorrelations among the cues. The investigation is still in progress, but some preliminary results will be used to illustrate the main principles of the proposed paradigm.

Methods

46 performances, three melodies, and three performers were included in the analysis. The melodies were Greensleeves, When the Saints, and Nobody Knows. The performances were made on the electric guitar. Four cues were analyzed: Tempo, loudness, articulation, and frequency spectrum. These were selected on the basis of studies showing that they (a) vary consistently as a function of the performer's expressive intention (Gabrielsson & Juslin, 1996; Juslin, 1997a), and (b) are used by listeners to decode the expression (Juslin, 1996b). In the performer models, the cues were the independent variables and the performer's expressive intention was the dependent variable. In the listener models, the cues were the independent variables and the mean rating on the adjective scale was the dependent variable.

Encoding experiments

Three professional guitarists (playing jazz, blues, and rock), aged 25-45 years, played each melody so as to communicate the following emotions to listeners: angry, happy, sad, and scared. They had to keep the pitches of the melody, but was otherwise free to vary all aspects of the performance to generate the prescribed emotional expression. The performances were first recorded, and then analyzed concerning the following acoustical parameters; (a) mean tempo, (b) mean articulation, (c) loudness (Leq), and (d) long-term average spectrum (LTAS). The performance data were subjected to one regression analysis for each emotion.

Decoding experiments

31 listeners, aged 21-52 years, rated each performance (1-10) on four adjective scales; angry, happy, sad, and scared. The majority of the listeners were musically trained, and the stimulus order was randomized for each listener. The mean ratings were subjected to one regression analysis for each emotion (for further information on the data collection; see Juslin, 1997a).

Results and Discussion

Table 1 presents the results from the regression analyses of the performer and the listener systems, in terms of the LME statistics, the multiple correlations, and the standardized beta weights for each of the cues. By excluding or including particular cases (i.e., performances) in the models, it was possible examine the effects of different melodies or performers on the relative cue weights of the systems. However, because of space limitations, only two types of models will be discussed here; one that includes all melodies and performers, and one that includes all melodies but only performer 1. Beginning with the LME statistics, it is clear that the performers were quite successful in communicating the emotions to listeners. Achievement ranged from .81 to .96 depending on the particular emotion and model. Notably, anger and happiness were better communicated than sadness and fear. The matching indices suggest that the performers' cue weights were better matched to the listeners cue weights in the cases of anger and happiness. That the achievement was as high as it was begs the question whether the communicative process can be improved at all. However, it should be noted that this study featured expert performers. Presumably, less accomplished performers would obtain lower degrees of achievement and be in greater need of improved cue utilization. All LME statistics were statistically significant (p<.05). The multiple correlations (M = .79) were typical of those obtained in judgment studies (.70 to .90). In other words, the linear models fitted the data well. This suggests that both the performers and the listeners used the cues rather consistently. One exception was the multiple correlation obtained for happiness (when all of the performers and the melodies were included in the model). This was largely because (a) the performers' use of
tempo differed across melodies, and (b) the happiness expression was more dependent on tempo than the other expressions. This was confirmed by a model (not shown in the table) including only performances of When the Saints by all performers: In this model, the multiple correlation (happy intention) increased to .85, and the ecological validity of tempo increased from .41 to .79. The relative weights in Table 1 should be interpreted as follows: Positive signs of the cue weights indicate, respectively for each cue, high tempo, high loudness, legato articulation, and much high-frequency energy in the spectrum. The great majority of the beta coefficients were statistically significant (p<.05). Interestingly, loudness and articulation had higher overall ecological validities than tempo and spectrum. The results generally replicate the data of our earlier studies (e.g., Gabrielsson, 1995; Gabrielsson & Juslin, 1996; Gabrielsson, Juslin, & Lindström, 1994; Juslin, 1996a; 1996b; 1997a; 1997b; 1997c). For example, both performers and listeners generally associated sadness with low tempo, loudness, legato articulation, and soft spectrum. However, the simultaneous description of the performers' and the listeners' cue utilization opens up the possibility of a direct comparison of the two systems. For each of the emotions communicated, it is possible to compare the relative weights given to each cue by performers and listeners, respectively. In those instances were discrepancies occur, this may be corrected either by changing the cue weights of the performer, or the cue weights of the listener, or both. For example, it is seen that the sad version by performer 1 should have been played softer, and that listeners attributed higher importance to tempo than performers regarding happy expressions. Low to moderate cue intercorrelations were found for tempo / loudness (r=.48), tempo / articulation (r=.08), tempo / spectrum (r=.14), loudness / spectrum (r=.42), loudness / articulation (r=.49), spectrum / articulation (r=.15). These intercorrelations reflect the ecological redundancy of the cues, which presumably helped the listeners to decode the emotional expressions.

Concluding Remarks
This study has concerned a situation in which encoding and decoding are made by integrating several uncertain but partly redundant cues into an unitary output. This type of task requires knowledge about relations among the variables, as envisaged by Brunswik (1956). That is, participants need information that allows them to compare the relations that they are using to the optimal relations for the task. This procedure is referred to as cognitive feedback (CBF). In contrast to outcome feedback (i.e., informing the participant about the correct answer), CBF has been found to improve the accuracy of judgments in many circumstances (Balzer, Doherty, & O'Connor, 1989). In a musical context, performers and listeners could be informed about how their relative weights compare to the optimal weights for successful communication. This information could be administered by means of cognitive feedback programs that include graphical presentation of regression models and LME statistics (cf. Cooksey, 1996).

What, then, is the answer to the question asked in the title of this paper? In light of the present results, the answer must be that performers can become better at communicating emotions to listeners by (a) having their performances (with different emotional expressions) recorded, analyzed, and described in terms of regression models, (b) letting these (idiographic) models be compared to some target population's (nomothetic) models, and (c) using cognitive feedback concerning how their cue weights compare to the optimal cue weights of the target population to improve cue utilization. Obviously, the present findings are limited in several ways: A larger number of performers and melodies (and perhaps cues) should preferably be entered into the regression analysis. However, this paper serves mainly to demonstrate the usefulness of the functionalist approach concerning the problem of how to improve the communicative process.

Acknowledgements: The research was supported by The Bank of Sweden Tercentenary Foundation through a grant to Alf Gabrielsson.
References


Juslin, P. N. (1997c). *Can the results from studies of perceived expression in musical performances be generalized across methods?* Manuscript submitted for publication.


Table 1: The main results from the regression analyses of the performer and listener systems.

(simultaneous regression models)

<table>
<thead>
<tr>
<th></th>
<th>Angry</th>
<th>Sad</th>
<th>Happy</th>
<th>Scared</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EMOTION</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PERFORMER-LISTENER RELATIONS (LME):</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(V) All melodies-All performers: Achievement ($r_a$):</td>
<td>.95</td>
<td>.74</td>
<td>.85</td>
<td>.81</td>
</tr>
<tr>
<td>Matching (G):</td>
<td>.99</td>
<td>.92</td>
<td>.94</td>
<td>.93</td>
</tr>
<tr>
<td>(VI) All melodies-Performer 1: Achievement ($r_a$):</td>
<td>.96</td>
<td>.92</td>
<td>.95</td>
<td>.83</td>
</tr>
<tr>
<td>Matching (G):</td>
<td>.99</td>
<td>.84</td>
<td>.94</td>
<td>.92</td>
</tr>
<tr>
<td><strong>PERFORMER MODELS (expr. intentions-cues)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(I) All melodies-All performers:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple Correlation R</td>
<td>.78</td>
<td>.81</td>
<td>.55</td>
<td>.81</td>
</tr>
<tr>
<td>Tempo</td>
<td>.11</td>
<td>-.26</td>
<td>.41</td>
<td>-.13</td>
</tr>
<tr>
<td><em>Ecological validities</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loudness</td>
<td>.47</td>
<td>-.46</td>
<td>.26</td>
<td>-.22</td>
</tr>
<tr>
<td>Articulation</td>
<td>.06</td>
<td>.73</td>
<td>-.16</td>
<td>-.63</td>
</tr>
<tr>
<td>Spectrum</td>
<td>.36</td>
<td>-.09</td>
<td>-.24</td>
<td>-.08</td>
</tr>
<tr>
<td>(II) All melodies-Performer 1:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple Correlation R</td>
<td>.84</td>
<td>.87</td>
<td>.73</td>
<td>.88</td>
</tr>
<tr>
<td>Tempo</td>
<td>.13</td>
<td>-.57</td>
<td>.47</td>
<td>.02</td>
</tr>
<tr>
<td><em>Ecological validities</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loudness</td>
<td>-.13</td>
<td>.23</td>
<td>.63</td>
<td>-.68</td>
</tr>
<tr>
<td>Articulation</td>
<td>.28</td>
<td>.52</td>
<td>-.37</td>
<td>-.43</td>
</tr>
<tr>
<td>Spectrum</td>
<td>.80</td>
<td>-.51</td>
<td>-.53</td>
<td>.13</td>
</tr>
<tr>
<td><strong>LISTENER MODELS (judgments-cues)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(III) All melodies-All performers:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple Correlation R</td>
<td>.73</td>
<td>.81</td>
<td>.65</td>
<td>.79</td>
</tr>
<tr>
<td>Tempo</td>
<td>.13</td>
<td>-.52</td>
<td>.57</td>
<td>.06</td>
</tr>
<tr>
<td><em>Functional validities</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loudness</td>
<td>.48</td>
<td>-.36</td>
<td>.18</td>
<td>-.58</td>
</tr>
<tr>
<td>Articulation</td>
<td>.05</td>
<td>.41</td>
<td>-.11</td>
<td>-.36</td>
</tr>
<tr>
<td>Spectrum</td>
<td>.30</td>
<td>-.06</td>
<td>-.10</td>
<td>.16</td>
</tr>
<tr>
<td>(IV) All melodies-Performer 1:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple Correlation R</td>
<td>.76</td>
<td>.86</td>
<td>.79</td>
<td>.91</td>
</tr>
<tr>
<td>Tempo</td>
<td>.13</td>
<td>-.58</td>
<td>.63</td>
<td>-.36</td>
</tr>
<tr>
<td><em>Functional validities</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loudness</td>
<td>-.07</td>
<td>-.31</td>
<td>.44</td>
<td>-.22</td>
</tr>
<tr>
<td>Articulation</td>
<td>.21</td>
<td>.30</td>
<td>-.15</td>
<td>-.47</td>
</tr>
<tr>
<td>Spectrum</td>
<td>.69</td>
<td>-.12</td>
<td>-.31</td>
<td>-.24</td>
</tr>
</tbody>
</table>
Expressive Performance in Jazz, Take 2: Ranges of expressive timing in jazz melodies

Richard Ashley
School of Music
Northwestern University
711 Elgin Road
Evanston, IL 60208-1200 USA
tel.: 847/491-5720
email: r-ashley@nwu.edu

Introduction
Earlier research (Ashley 1996, 1997) has shown that jazz musicians performing the melodies of standard ballads (such as 'My Funny Valentine' and 'My Foolish Heart') follow a relatively well-defined set of procedures for transforming the melody's notated rhythm in order to make it less 'rigid' and more 'expressive.' There is a kind of jazz composition closely related to the 'standard' ballad, and yet quite distinct from it; this is the newly-composed ballad, frequently ensconced in a single, authoritative recording. Such compositions often display a melody similar in style and structure to the standard ballad, and yet these melodies are often performed with less rhythmic freedom than is frequently found with performances of standards. As such, the performance practice surrounding such compositions acts as a kind of test of the hypotheses and observations made with regard to expressive timing in the performance of standards. The primary question here is to what degree these more straightforward melodic treatments exhibit the same structural features and performance-practice tendencies found in renditions of standards.

Analysis of recorded examples (Ashley 1996) found that 'standard' ballads were performed in a number of rather regular ways. The traits displayed by standard ballad performances included:
• the underlying beat and meter provided by the rhythm section remained very stable for the most part;
• the soloist had considerable freedom to displace the attack-points of melodic tones, either earlier or later, over this steady background;
• either a 'start late and speed up' or 'start early and slow down' strategy tends to predominate for a given phrase of the melody;
• within these overall rubato strategies, cadential points tend to be displaced relatively little ('cadential anchoring');
• downbeats are handled differently than other locations; specifically, at times downbeats are 'anchored' in place like mini-cadences, with little offset from the underlying metric grid;
• rhythmic alteration tends to preserve important proportional rhythmic figures; that is to say, if a motive has three notes in the ratio 1:1:1 in the original, the expressive transformation of this figure will tend to preserve the original ratio, and thus preserve the character of the motive;
• finally, melodic notes belonging to the underlying harmonies are often displaced to a much greater extent than are nonharmonic tones, especially harmonic tones coming on a downbeat (the hypothesis here is that displacing harmonic tones results in a greater degree of melodic tension and thus a higher degree of expressive potential).

In summary, the expressive transformation of a jazz ballad melody through altering its rhythmic aspects reveals an intimate relationship to musical structure at a number of different
levels. This is reflective of many earlier studies, showing the relationship of rubato to meter and harmony (cf. Bengtsson & Gabrielsson 1983, Todd 1985, Desain & Honing 1992, Repp 1992, Palmer 1996, and many others).

**Materials and methods**

The sources of data for this paper are commercially-available recordings of jazz ballads, specifically ‘Naima’ by John Coltrane and ‘Good-bye, Pork-Pie Hat’ by Charles Mingus. These are compositions which are extremely well-known in the jazz community and have, in the four decades since they were first composed and recorded, been widely accepted, played, and recorded by others. These compositions are ballads—slow, lyrical melodies—but are not ‘standards’ in the manner of songs like ‘My Funny Valentine.’ The methodology of this study is the same as that used in Ashley (1996); a summary is given here.

The first step in analyzing the performances of these compositions, after selecting the compositions themselves and recordings of them, is to make digital soundfiles of the CDs. This was done on Macintosh computers, using direct digital transfers from the CDs to 16-bit soundfiles at a sampling rate of 44.1 kHz, to ensure maximum fidelity to the original. After producing these digital soundfiles, the files are ‘marked’ to show the points in time at which the notes which will be studied begin. In this, two sets of notes are used: the melodic tones played by the soloist(s), and notes from the rhythm section (typically bass or drums) which give the underlying metric ‘grid’ against which the melody is performed. The onset of these notes is found by visual and aural examination of the soundfile—a time-consuming process, but one which is relatively precise given sufficient experience. The beginning of the note is determined by a rule: the note onset is defined to be the amplitude zero-crossing most immediately preceding the beginning of the (quasi)periodic part of the note for pitched instruments or that zero-crossing most immediately preceding the beginning of the ‘noise’ portion of the sound for a cymbal note.

In this style it is typical for expressive performance of melodies to involve a soloist altering the ‘nominal’ or notated rhythm of the composition over a relatively steady and unvarying background provided by the rhythm section (typically piano, bass, and drums). This performance practice thereby differs significantly from much of the music typically used in studying expressive performance (for example, Mozart piano sonatas) in that the tempo and timing of the accompaniment may not co-vary with the melody; rather, the ‘rubato’ used in this style is more like that of earlier music (even that attributed to Chopin—cf. Hudson 1994).

The note onsets obtained by careful examination of the soundfiles are then converted to millisecond placements within the file. A metric grid, representing the attack-point of each beat in the recording, is produced first. This is followed by a ‘nominal’ representation of the melody; that is, given the metric grid as performed, and the notated version of the melody, where would the onsets or attack-points of the melodic notes fall if played exactly as notated against the metric grid? This ‘nominal’ representation of the melody’s attack-point rhythm is then used as a standard against which the actual melodic performance is compared. The most critical aspect of this comparison is **attack-point displacement (AD)**. A melody note which is performed earlier than would be predicted by the ‘nominal’ melodic rhythm has a negative AD; one which is performed later than in the ‘nominal’ rhythm has a positive AD.

**An example: John Coltrane’s ‘Naima’**

John Coltrane’s ballad ‘Naima,’ written in honor of his wife, is a much-admired example of the original jazz ballad. Coltrane himself recorded the composition twice, in March and December 1959. The later of these two versions or ‘takes’ was for years the only one
available (the ‘main’ take); the earlier, ‘alternate’ take, is also now available with the main take on the Giant Steps CD (Atlantic 1311-2). These recordings are remarkable not only for the composition but also for what they show about Coltrane’s expressive strategies. Due to space limitations, this discussion will focus only on the ‘A’ section—the first two phrases of each performance, more extended data and discussion of this composition and others are available from the author. The notation for the first phrase is given below (Fig. 1); the second phrase is a repetition of the first.

![Notation for the first phrase of 'Naima']

Figure 1: ‘Naima’, John Coltrane; ‘A’ section (mm. 1-8)

The two performances are in the same key, have the same instrumentation (tenor saxophone, piano, bass, and drums—although with different personnel for the two versions), and are roughly in the same tempo (main take: quarter note avg. = 1000 msec; alternate take: quarter note avg. = 1005 msec). Nevertheless, one striking difference in the main melodic section emerges between the two versions. In the earlier, alternate take, the drummer consistently plays the hi-hat cymbal with his left foot on beats 2 and 4 of a 4/4 bar, providing a very strong metric grid; this is omitted in the later version.

When looking at the ADs for the two different performances of this melody (Fig. 2), a few main points emerge quickly. These are:

![Graph showing attack-point displacements (ADs) in two performances of 'Naima']

Figure 2: Attack-point displacements (ADs) in two performances of ‘Naima’
compared to performances of 'standard' ballads, both versions use relatively small ADs from the 'nominal' melodic positions; the largest deviations are only about 0.2 beats, as opposed to the 2 or more beats often found for ADs in 'standard' ballads;

both takes are very similar with regard to the mean and standard deviation of ADs (main take: 88/54 msec; alternate take: 84/47 msec)

the main take involves ADs which are both early (negative) and late (positive); the earlier version utilizes only ADs which are positive (late);

the absolute range of ADs used in the main take is 338 msec. as opposed to 171 msec for the alternate take, a difference of about 2:1;

the main take makes use of a relatively clear 'start late-speed up' strategy in both phrases, although the second phrase is more extreme than the first, whereas the alternate take lacks the kind of clarity of shape found in the other version;

the main take is, in its clear shape, more like the performance of a standard ballad than is the alternate take.

A more detailed look at the structure of the performance and its relationship to harmony and meter (Fig. 3) reveals the following characteristics:

![Figure 3: Attack-point displacements (ADs) in two performances of 'Naima,' with regard to structural aspects](image)

- in the alternate, earlier take, the mean ADs are about the same for downbeats and other beats (79/88 msec), and for melody notes belonging to or outside the underlying harmonies (86/82 msec);
- in the later, main take, the mean ADs are differentiated in a relatively small ways between downbeats and other beats (37/54 msec) and much more between harmonic and nonharmonic tones (69/26 msec).

In sum, the later version, the one Coltrane chose to release commercially, shows a greater sensitivity to meter and harmony, in that more 'stable' musical events (downbeats and
harmonic tones) are displaced more than their less stable counterparts—in part, it seems, to provide greater musical tension in the performance than would otherwise be present.

**Conclusion**

The jazz listener might note that newly-composed jazz ballads such as 'Naima' tend to be played 'straighter' than standards—that is, with less rhythmic freedom. Yet, one is struck by the fact that these performances are highly expressive. This study indicates that the reasons for this lie not in basic differences between performance strategies in the two types of compositions, but in the depth or degree to which these strategies are applied. If one thinks of expressive timing as in some sense a functional variation upon a strict tempo (the traditional 'tempo curve,'), the differences between these repertoires may be thought of not as a difference in the kind of functions used, but in their relative amplitudes. The 'standard' ballad will use a function with a greater amplitude; the jazz ballad, one with a lesser amplitude. In this way, one finds a corollary to a finding in the performance of 'classical' music, that when a performer tries to play 'non-espressivo,' the main effect is the diminishing of his natural expressive tendencies (Bengtsson & Gabrielson 1983).

The jazz musician is highly skilled and possess great stores of knowledge about musical structure which can be called upon quickly, not only in 'improvisation' but also in the more routine aspects of musical expression, as in playing an unadorned melody. The thoughtfulness with which these master craftsmen ply their trade is extraordinary (cf. Berliner 1996), and the present study goes to show how even the subtle aspects of the craft reflects this deep comprehension and mastery of the structuring potentials of the idiom. Far from being 'off the top of the head,' these musicians' response to structure as shown in their habits of melodic performance demonstrates the depth of skill they have attained.

**References**


ANALYZING JAZZ RHYTHMS:
THE MICROSTRUCTURE OF WALKING BASS LINES

BERNHARD GÁL

University of Vienna (Austria)
bernhard.gal@blackbox.at

Introduction

This article focuses on the analysis of one of the fundamental ingredients of most jazz performances, the so-called walking bass accompaniment. In conventional jazz performance walking bass lines are coherent, mainly improvised jazz phrases, which serve both as the harmonic basis and carry the beat, i.e. the underlying pulse. The steadiness of the beat has to be differentiated from the machine-like monotony of a rhythm computer or metronome. Moreover, systematic deviations of the various parameters have been observed.

Method

(Data collection - Recordings)

Five double bass players participated in this experiment, each of them performing two walking bass lines read from notated transcriptions, played at two different tempos, thus resulting in 20 individual interpretations altogether. The recordings were made in mono using an AKG C5600 condensator microphone and went directly to R-DAT (Sony DTC-59ES). To facilitate the succeeding analysis, the musicians had been instructed to follow a click track (i.e. pre-recorded metronome clicks), which was also recorded on the second channel of the R-DAT tape.

(Analysis)

The analysis of the 20 walking bass interpretations comprises the measurement of tone onset / offset and intensity values. The digital recordings were transferred to S_Tools (a music and speech analysis workstation developed at the Acoustics Research Laboratory of the Austrian Academy of Sciences). Using a waveform- and RMS-contour editor the sound files were segmented into signal frames of one bar length, which was necessary to enlarge the resolution for the following measurements. The analysis was based on these segments transformed into RMS-values and performed by means of graphic and acoustic editors. (Fig. 1) The metronome clicks of the second channel served as a reference for the tone onset measurements. Intensity values were taken at the amplitude peaks of the respective tones. The accuracy of the measurements varied between 1-3 milliseconds, depending on the respective tempo of the sound examples. All data was fed into a statistic program for subsequent descriptive statistics.
Arithmetic means and variances were used in order to determine the range of timing, duration and intensity changes. The statistical evaluation showed several tendencies (as described below) that were used to create systematically modified versions of the two walking bass transcriptions. The starting point was to transfer the data of the notated duration and interval values into a MIDI-Sequencing system (Steinberg Cubase 2.0). Then several different examples were synthesized under systematic variation of timing (tone onsets), accentuation (tone duration $D_{on}$) and intensity respectively (Fig. 2). The MIDI-sequencer controlled a Kurzweil K2000 Sampler, where double bass samples for each pitch could be addressed. As a matter of fact, the interaction of software resolution, tempo and the K2000 allowed modifications of the time structure to steps of 2.6 milliseconds. Intensity could be controlled in 0.16 dB steps.

(Listening experiments)

Ten subjects participated in the experiment, grouped as experienced and inexperienced listeners. The purpose of the test was to evaluate the impact of systematic variations on the subjective perception of jazz rhythm. The participants were asked to rank the sound examples, always presented in groups of three, with regard to their perceived „jazziness“ of the examples. No further instructions were given. The aim was to receive intuitive and emotional rhythm responses and to avoid analytically based answers. They would possibly have shown the degree of musical knowledge and ability of the participants only. The whole test lasted twenty minutes and was repeated after a short break with newly randomized order of examples.

Results

It has been assumed that the steadiness of his playing, i.e. the ability to maintain a steady pulse is one of the basic criterions for the quality of a jazz musician. This ability seems especially important in the performance of walking bass lines, carrying the beat and therefore being the reference point for all other participants of a jazz performance. The recent analysis

---

1 For an elaborate presentation of the analysis-by-synthesis approach see Bengtsson and Gabrielsson (1983).
of the four walking bass examples revealed small, but perceptual relevant deviations from the notation compared to similar studies. (e.g. Reinholdsson, 1987). The following general tendencies were found in these deviations.

**General tendencies (referring to one idealized bar)**

**Timing:** (Fig. 3) Range of most commonly occurring deviations: -50 ms to +50 ms
Tendency: Retardation of the beginning and subsequent acceleration. (more distinct at 180 bpm)

**Duration $D_{ii}$:** (Fig. 4) Range of most commonly occurring deviations: 93-106%
Tendency: Shortening of the first three beats and lengthening of the fourth beat. Of course, these results show the correlation between timing and duration.

**Intensity:** (Fig. 5) Range of most commonly occurring deviations: -20 to -5 dB
Tendency: Accentuation of the second beat and slight diminution of the first and fourth beat.

**Offbeats:** Offbeats and their respective neighboring beats were excluded in this article as they show different tendencies, whose description would exceed the range of this contribution.

---

The listening experiment revealed two results:

1. Experienced listeners showed a preference for the intensity variations (48.4%), followed by the mechanical version (40.6%). On the other hand, timing and duration patterns were ranked lower in „jazziness“ (6.2% / 4.7 %). A necessity to introduce more elaborate performance rules, including the analysis of higher structural levels, has to be concluded.
2. Inexperienced listeners marginally preferred the intensity patterns, too (31.2%), but didn’t show further preferences. This result leads to the assumption that certain difficulties in perceiving the parameter variations occurred and consequently ranking was complicated.

**Conclusion**

The current results are preliminary and show general trends only. Deviations measured from live performances cover a much wider range than the observed statistical tendencies. Striking parallels between different performances of one piece, particularly when played by the same musician have been revealed. A closer look at this phenomenon might be a rewarding task for further research.

My findings have to be considered as a survey study of this subject. Some questions remain open for future research:

- How do factors like personal style, mood, environment, interactions with other musicians, emotional characteristics etc. take influence on performances?

- How can the appearance of specific phenomena like offbeats, ghost notes etc. be integrated in the analysis?

- How can systematic deviations at higher levels be described adequately? (e.g. comparing different periods (=choruses) of one piece with regard to structural, harmonic and motional relations)

**References**


Music Analysis: Modelling
1. Introduction

The ability to infer beat and metre from music is one of the basic activities of musical cognition. It is a fast process: after having heard only a short fraction of music we are able to develop a sense of beat and metre and tap our foot along it. Even if the music rhythmically complex, containing a range of different time values and probably syncopation, we are capable of inferring the different periodicities of it and synchronising to them. A rhythmical sequence usually evokes a number of different pulse sensations, each of which has a different perceptual salience. Furthermore, for a given piece of music, the most salient pulse sensation can vary between listeners. According to experimental literature (cf. Parncutt, 1994), the range of most salient pulse sensations lies between 67-150 events per minute, corresponding to periods of 400-900 ms, the greatest salience being in the vicinity 600 ms.

In addition to the basic pulse level, we simultaneously perceive periodicities on higher hierarchical levels (Palmer & Krumhansl, 1990); the longer periodicities are integral multiples of the periodicity of the basic pulse. This results to a percept of periodically alternating strong and weak beats, corresponding to the generally accepted definition of metre (Lerdahl & Jackendoff, 1983).

Parncutt (1994) distinguishes between two forms of processing of temporal information: serial and periodic grouping. Serial grouping refers to the processing of temporally adjacent events, including motives, themes, phrases etc. (Lerdahl & Jackendoff, 1983). Periodic grouping, on the other hand, refers to the processing of nonadjacent events. According to Parncutt (1994), periodic grouping can be divided into two stages, namely pulse sensation and perceived metre.

The theory of Lerdahl and Jackendoff (1983) distinguishes between three kinds of accent contributing to the rhythmic organisation of music: phenomenal, structural, and metrical. They rely, respectively, on sensory, structural, and schematic sources of evidence (Palmer & Krumhansl, 1990). Phenomenal and structural accents serve as perceptual input to metrical accents (Lerdahl & Jackendoff, 1983).

Attempts to model the perception of pulse and metre have relied on a diversity computational formalisms. These include, among others, rule-based systems (e.g., Longuet-Higgins & Lee, 1982), statistical approaches (Palmer & Krumhansl, 1990; Brown, 1993), optimisation approaches (Parncutt, 1994), and connectionist models (Scarborough, Miller & Jones, 1992). A common feature of these models is that they deal with idealised rhythms, i.e., rhythms comprised of precise durations such as those found in a musical score. Models that attempt to find rhythmic parsing directly from performances have been proposed by Chung (1989),

This paper introduces a metaphorical model of how the sensation of metre builds up when listening to music. The model uses as input accoustical representations of real musical performances. At the first stage, an event detector generates a sequence of discrete impulses, where each impulse represents an event (e.g. onset) in the input stream. Beat perception is modelled by means of a bank of competing adaptive oscillators, modified from [Large & Kolen, 1994]. By continuously adapting their phase and period, these oscillators attempt to synchronise to periodic components of the sequence of pulses. They can follow local variations of tempo and produce expectations of next beats.

Metre is extracted by means of competing subharmonic oscillators, each of which represents a possible metrical interpretation, or schema. The degree of resonance of each subharmonic oscillator is built up by leaky integrating the coincidence factor of the last pulse, i.e. the degree to which the last pulse coincided with the instant of firing of the oscillator. This approach takes automatically into account the durational accents of events and the recency effect (Parnutt 1994). The perceived metre is defined to be the metre represented by the winner, i.e., the subharmonic oscillator with the highest resonance.

2. Model

The structure of the model is schematically depicted in Figure 1.

![Fig. 1. A schematic overview of the model.](image)

At the present stage, the input to the model consists of monophonic melodies. The sequence of impulses is generated by an onset detector, which calculates FFT power spectra every 20 ms, using a 40 ms Hamming window. Let $F(t)$ denote the Fourier power spectrum at time $t$. The preprocessing module generates a pulse, if

$$
\begin{cases}
\Phi(t) > \delta \\
\Phi(t - \Delta t) < \delta
\end{cases}
$$

(1)

where $\Phi(t)$ is defined by
\( \Phi(t) = \| F(t) - F(t - \Delta t) \| + \left( \frac{1 - \frac{F(t) \cdot F(t - \Delta t)}{\| F(t) \| \| F(t - \Delta t) \|}}{\| F(t) \| \| F(t - \Delta t) \|} \right) \), \hspace{1cm} (2)

and \( \delta \) is a threshold value.

The perception of pulse is modelled by a bank of competing adaptive oscillators working in parallel. The state of an oscillator at a given time, \( t \), is defined by its phase \( \phi \) and period \( p \). The two latter are related by

\[ \dot{\phi}(t) = \frac{1}{p(t)}, \hspace{1cm} (3) \]

where \( \dot{\phi}(t) \) stands for the time derivative of \( \phi \) at time \( t \); \( \phi \) can be constrained to stay within the range \(-1/2 \leq \phi < 1/2\) by means of a modulo operator.

The output of the oscillator at a given time, \( t \), depends on its phase \( \phi \) according to

\[ o(t) = 1 + \tanh[\gamma(\cos 2\pi \phi(t) - 1)]. \hspace{1cm} (4) \]

The oscillator thus generates a pulse whenever \( \phi \geq 0 \); the temporal width of the pulse can be controlled by the gain parameter \( \gamma \).

The oscillator entrains only to input impulses which occur within these output pulses. When such an impulse occurs before the oscillator expects it to occur (\( \phi < 0 \)), the phase is increased and the period is shortened, and vice versa. The amount of adaptation of both phase and period is proportional to the first time-derivative of Equation 4 (see Large & Kolen, 1994).

While the adaptation of the oscillator of Large and Kolen is defined by delta rules (difference equations), the present approach utilises continuous adaptation of phase and period (differential equations). This approach, we believe, makes the oscillator more probable to synchronise to those periodic components of the impulse sequence which contain strong durational accents, i.e., are followed by long IOIs.

Let the impulses occur at times \( t_i, i = 0,1,2,... \). At each \( t_i \) we define the subsequent adjustment of phase and period to be proportional to the first time derivative of Equation 4 (see Large & Kolen, 1994):

\[ \Delta(t_i) = -\frac{1}{2\pi} \frac{\partial o(t_i)}{\partial t} = \text{sech}^2[\gamma(\cos 2\pi \phi(t_i) - 1)] \sin 2\pi \phi(t_i), \hspace{1cm} (5) \]

Furthermore, we define a gating function which adjusts the rate of adaptation after the last impulse:

\[ \Gamma(t) = 256 \varphi(t)^2 e^{-\varphi(t)}, \hspace{1cm} (6) \]

where \( \varphi \) is the amount of phase increase since the last impulse,

\[ \varphi(t) = \phi(t) - \phi(t^*), t^* = \max\{t_i; t_i < t\}; \hspace{1cm} (7) \]

the gating function \( \Gamma \) is normalised according to \( \int_0^\infty \Gamma(t) dt = 1 \). The time evolution of phase and period are now defined by equations

\[ \begin{cases} \dot{\phi}(t) = \frac{1}{p(t)} - \eta_1 \Gamma(t) \Delta(t^*) \\ \dot{p}(t) = \eta_2 \Gamma(t) \Delta(t^*) p(t) \end{cases}, \hspace{1cm} (8) \]
where \( \eta_1 \) and \( \eta_2 \) are the coupling coefficients for phase and period, respectively. The gating function \( \Gamma \) is defined so that onsets followed by an interonset interval (IOI) of less than about 100 ms do not give rise to any significant adaptation; this is concordant with the fact that the minimum discriminable IOI is of the order 50-100 ms (Parncutt, 1994). For onsets followed by longer IOIs, most of the adaptation occurs within the first half cycle after the input pulse.

An oscillator described above is capable of synchronising to the beat of a complex rhythm only if its initial period is sufficiently close to that of the beat. Therefore, pulse detection is carried out by means of a bank of such oscillators working in parallel, each of which has a different initial period; the latter are set so that they cover the range of most salient pulse sensations. The degree of resonance of each oscillator, \( \rho \), indicates how well it has succeeded in synchronising with the beat; its time evolution is defined by the equation

\[
\dot{\rho}(t) = \sigma(t\ast)\Psi(t) - \frac{\rho(t)}{kp(t\ast)},
\]

where the gating function \( \Psi \) is defined by

\[
\Psi(t) = 32(t - t\ast)e^{-4(t - t\ast)}.
\]

According to Equations 9-10, if an impulse occurs during the output pulse of an oscillator, i.e., when \( \sigma(t\ast) = 1 \), the degree of resonance of the latter starts to increase until the next impulse; the gating function \( \Psi \) is defined so that its integral approximates the dependence of phenomenal accent on interonset interval (Parncutt, 1994). When no excitation is present, the degree of resonance decays with a time-constant \( kp(t\ast) \); the somewhat arbitrarily chosen value \( k = 5 \) was used in the experiments. At any given instant of time, the oscillator with the highest degree of resonance is defined to be the system's interpretation of the beat. Figure 2 presents an example of the adaptation of two oscillators having different initial periods.

![Figure 2](image)

**Fig. 2.** Adaptation of two oscillators to a rhythmic sequence of alternating sixteenth- and dotted eighth notes. The beat period is 480 ms; the initial periods of oscillator 1 and 2 are 510 ms and 450 ms, respectively; \( \gamma = 1.5, \ \eta_1 = 0.4, \ \eta_2 = 0.2 \) (see Eqs. 4 and 8). Panel a displays the input pulse train and the output of each oscillator; panel b the period in seconds as a function of time, and panel c the degree of resonance as a function of time. Oscillator 1 synchronises with the onsets of the dotted eighth notes, which leads to a high value of resonance.

To model the perception of metre, each adaptive oscillator is coupled to a pool of subharmonic oscillators whose periods are integral multiples of that of the adaptive oscillator. Each subharmonic oscillator is characterised by the tuplet \( (P, \phi_0) \), where \( P \) denotes the ratio of the
its period and that of the adaptive oscillator, and $\phi_0$ its initial phase. In the present model, subharmonic oscillators $(2, 0)$, $(2, 1/2)$, $(3, 0)$, $(3, 1/3)$, and $(3, 2/3)$ are used, corresponding to all the binary and ternary interpretations of metre.

![Diagram of oscillators](image)

**Fig. 3.** The outputs of an adaptive oscillator and its five subharmonic oscillators.

The degree of resonance of each subharmonic oscillator is calculated similarly to that of the adaptive oscillators, that is, using Equations 9-10. At any given instant of time, the subharmonic oscillator with the highest degree of resonance is defined to be the system's interpretation of the strong beats.

### 3. Observations about the experiments

The preliminary experiments carried out show that the model is capable of extracting pulse and metre from the melodies it is presented with, assuming that the latter are rhythmically sufficiently simple, i.e., that they do not contain extensive syncopation or abrupt tempo changes. The leaky temporal integration mechanism utilised in calculating the degree of resonance of the subharmonic oscillators allows the system to detect changes of metre in real time; an example of that is depicted in Figure 4.

The model works on-line with temporally local information. Therefore, in addition to analysing the metre of the input, it is also capable of synthesising, i.e., tapping along the music.

### 4. Conclusion

Similarly to other models of pulse and metre perception published till now, the present version of the model derives phenomenal accents from the durations of tones only. While durational accent seems to be the greatest contributor to phenomenal accent (Palmer, 1989), the latter can also be evoked by changes in loudness, timbre, melodic contour, and implied harmony (cf. Parnacutt, 1994). In complex musical rhythms, the IOI structure very often does not contain enough cues for the definition of the underlying metre. Future research will focus on how the contributions to phenomenal accent mentioned above could be included in the model.
Fig. 4. Detection of change of metre in a musical performance. The graph displays the degree of resonance of two subharmonic oscillators, \((2, 0)\) and \((3, 2/3)\). The melody used as input starts in 4/4 meter and switches to 3/4 meter at about 7 seconds. About 2 seconds after the change, the degree of resonance of subharmonic oscillator \((3, 2/3)\) grows beyond that of oscillator \((2, 0)\), indicating that the system has detected the change of metre.

References


MODELLING MUSICAL CONVENTIONS

Ulf Berggren, Ph.D., Uppsala

During the era from approximately 1750 through 1800, both composers of genius and less skilled composers extensively worked with stereotypical constructs, or conventions. One reason to use conventions was to facilitate the listener's comprehension of the composer's ideas. Some conventions regulated musical form, for example "sonata form", a conceptual framework used by numerous composers. Other constructs commonly used in composition regulated the making of phrase structure, harmonic progressions, rhythmic progressions, dissonance treatment, etc. I would like to point at an example of highly conventional music: a harpsichord and violin sonata written by the young Wolfgang Amadeus Mozart, sonata in C major, K. 6. The first movement abounds in stereotypes, such as adherence to formal schemes of the time (interpunctive form); Alberti bass; much repetition of melodic formulas, etc.

Expressed in the terminology of eighteenth century music theory, both the amateur and the professional composer must know of a sufficiently large set of mechanical rules of composition. Such rules were described in the eighteenth century didactic theory by, among others, Heinrich Christoph Koch. In his Versuch (1782-93), he equals the use of mechanical rules with common practice. He believed that mechanical rules govern a significant part of composition, a part that does not require esthetic insights.

In a similar sense, the term mechanical is used nowadays in definitions of algorithms. An algorithm is a mechanical procedure which may function entirely without human intervention. Roger Penrose (1994) argues for the existence of a genuine dualism between algorithmic and non-algorithmic thinking. In his view, there exists considerable algorithmic work of the brain which is unconscious, while there is non-algorithmic work which is conscious. Following this distinction, one may assume there are two cognitive levels in composition of music: one algorithmic level and one esthetic. To make valuable classic music, insights in the esthetics of the art are necessary. But in order to compose "correct", though simple, classic music there is no need for genuine esthetic insights, standard compositional tools are sufficient. While, in practice, a composer always puts at least some esthetic insight in his work, it is fully realistic to perform composition completely algorithmically: the algorithmical part of composition may cover all obligatory phases of the compositional process, such as the making of phrase structure, harmonic progressions, melody, accompaniment, etc.

A composer's making of non-trivial tonal music presupposes complex mental activity. An effective way to achieve deeper understanding of composition as an activity is to apply methods from general systems theory and to formulate models of the composition process. Then the musical knowledge (or competence) required to compose a "correct", but not necessarily esthetically good, piece of music is formalized. Normally, the cognitive process in making a musical work involves a
large set of decisions. Such decisions can be modelled by computational means: in terms of a search space one has to determine those choices generating paths that result in "correct" musical pieces in accordance with the classical norm. One obvious output from models that simulate the composition process is musical samples, possibly new musical pieces, in a certain style.

A generative model for classical composition

I will now describe my own approach to the problem of modelling music conventions used in composition. The objective of my dissertation, *Ars Combinatoria* (1995), is to make a model that simulates composition entirely automatically. Conventional material of many kinds has been derived from the first movements of piano sonatas by W. A. Mozart. Simulations of piano pieces, approximately 100 bars in length, have been effectuated. To make it possible to generate a complete sonata movement, a substantial set of conventions has been formalized and structured in terms of a programming language (Prolog). The model is quite complex and consists of three submodels, each of which includes numerous processes that operate on compositional units of different kinds. The formalism of the model consists of conditional statements (Prolog rules). Some of these statements have a rather obvious cognitive counterpart (e.g. compositional rules), while others handle more technical processual aspects with less immediate musical significance.

Representation of compositional units

Essential to the proposed modelling of composition techniques is the use of symbolic representations of musical knowledge. Compositional units are extensively stored and processed as symbols and schemes of symbols (lists). A symbol may denote a voice type, such as "alberti(1/16)" standing for the voice type Alberti bass in sixteenth notes. The standard length of such units is stipulated to be one quarter note. Other symbols denote harmonic functions. The naming of harmonic functions in the model corresponds to commonly used terms; e. g., a tonic is assigned the symbol "t", while a dominant seventh chord is assigned the symbol "d7". Schemes of harmonic functions include one symbol for each quarter unit. Thus, if the tonic function is to remain for one bar in common time, the data representation is the symbolic scheme (or list) [t,t,t,t]. Here, one symbol is assigned to each quarter note unit. A list appropriate to the two first bars of a sonata form movement may be [t,t,t,d,d,d,d].

A compositional unit on a higher level in the hierarchical structure corresponding to a musical piece includes units from lower levels. On the highest level, formal parts are assigned symbols such as "p1" denoting the first main period, that is the exposition. On the next structural level, formal parts are denoted with symbols such as "th1", which stands for the first group. On yet another level, the cadence is an example of a concept which includes units of harmonic, rhythmic, and melodic categories. The construction of cadences leads to the selection among different sorts of harmonic schemes (e.g. a half cadence is assigned a harmonic list that
ends with a dominant chord) and among duration schemes, as well as other choices.

In much classic instrumental music the play of texture types is structurally significant. In contrast to many instrumental movements from the baroque period, in a classical piece texture type may frequently change from one bar to another. Change of texture is one of the criteria that have been used to define a particularly important kind of unit in the model, the building block. A building block is a composite unit including several smaller units on different abstraction levels, finally actual notes. The range of a building block is defined typically to be one or two bars. Within a building block there are in some special cases only one, mostly two or more voices related by voice leading principles. For instance, the texture type of a building block may consist of two high voices coordinated by a complementary rhythm, plus a low voice.

It should be added that, while symbolical representation is essential in the model, some categories of musical units are stored numerically, e.g. rhythmical units ("duration groups", which complement the symbolic representation of rhythms), and successions of melodic intervals ("interval groups").

**Compositional rules**

Compositional rules make a very important category of conventions. Traditionally, composition of music in classical style is founded on the use of compositional rules, rules that regulate voice leading, doublings of notes within chords, and many other things. The making of a sonata movement follows the various rules that have been codified by, among others, Koch. There seems to be much consensus about the application of such rules, not only in didactical circumstances, but also in "serious" composition during the classic era.

The third submodel (Comp) includes a relatively large set of compositional rules corresponding to the rules discussed by Koch and later theorists. These rules control each voice within a building block with respect to notes tried by the generative mechanism. The categories of rules that check the writing include well-known rules for dissonance treatment, among them rules that control how characteristic dissonances (the minor seventh etc.) are resolved. Other rules accept certain melodic curvatures, and prohibit against unwanted curvatures such as "pentatonic" figures. There are rules that control doublings of chord tones, and prohibit against the doubling of leading tones, e.g. the third of major dominant chords. Other categories of rules prohibit against parallel fifths and octaves. Moreover, there are rules that check the textural writing when voices are coordinated in thirds or sixths, or when voices complement each other rhythmically. Yet another category contains rules that regulate certain melody notes in relation to harmonic functions: for example, within composition of classic music (and much other tonal music), there is a rule that says that the final strong cadence of a piece should end on the tonic.
The generative process

Within the simulation process there is a large number of steps between the initial steps (selection of musical form) and the final ones (production of notes in a score). Compositional units are selected from databases where they are stored, and combined by different processes in stylistically appropriate ways. Each of the numerous Prolog rules of the generative process consists of a set of conditions (possibly only one) and a consequence. The consequence of a Prolog rule often means that a tried compositional unit is accepted for further processing.

The first submodel, called Generate Goal, generates form, large periods, and the principal structure of the building blocks of the entire sonata movement. The second and third submodels (Plan resp. Comp) generate the content of one building block at a time. Schemes are selected matching the conditions for the building block to be composed: harmonic scheme, voice type schemes (one for each voice), duration schemes, and principal melodic movement lists (or schemes of motions between interval groups). Finally (in the third submodel), a detailed structure (musical notes) is generated for one building block at a time.

The selection of compositional units has to match the constraints of a given context. The conditions specify metric position within the bar, voice type (characteristics for melody, bass, and inner voices), and some other things. For example, to construct an Alberti bass figure of four notes, appropriate compositional units have to be selected: interval groups (which can be fifth up, third down, third up, fifth down), and duration groups (containing eighths or sixteenths). Subsequently, the compositional units are tested in order to match harmonical and other constraints. In some instances, the selection of compositional units involves randomizing. For example, at the beginning of a sonata movement the succession of note durations in a voice is selected with some randomizing from a set of duration schemes that match the condition "start of movement".

In sum, the mechanical nature of conventions in instrumental music of the classical era makes them well suited for computer simulations. More specifically, I have pointed at two categories that have proved to be particularly useful in modelling music conventions, symbolically encoded compositional units and compositional rules.

References


Machine Versus Human: Responding to the Task of Identifying Eras for Selected Keyboard Pieces

Besim Vefa Atalay, MSAI, Artificial Intelligence Center
Robert Placek, Professor of Music, School of Music
The University of Georgia
Athens, GA 30602, USA.

batalay@bach.cba.uga.edu, rplacek@uga.cc.uga.edu

Abstract

The musical era to which a certain piece belongs can usually be recognized by most trained musicians without much effort. However, it is an abstract categorization problem due to the overlapping characteristics of different musical eras and to the difficulty in strictly defining them. This seriously limits what can be achieved by formal analysis alone. The goal of the research effort described in this paper is to develop a method to identify the musical era of compositions using ANNs on a computer, and compare the results to a similar classification experiment on human subjects not having a musical background. This paper compares and contrasts two different sets of experiments in this field: Firstly, a computer implementation using a method of extracting relevant data from printed sheet music for training an artificial neural network is described. With the developments in the area of connectionism, there has recently been successful use of Artificial Neural Networks (ANNs) for music related problems like pitch and tonal recognition, and to a lesser extent, music composition. It is demonstrated that ANNs could also be applied to higher level cognitive music problems like the recognition of musical style, for which the computer-related work done so far has been non-connectionist. Secondly, a listening experiment is performed on undergraduate students taking a Music Appreciation class. It is shown that within the given framework, an accuracy rate of around 83% on the computer implementation is achievable on a three-way test to distinguish among the Baroque, Classical and Romantic eras, as opposed to the 57% average on untrained human subjects. It is also shown that human subjects take different criteria than the computer implementation in classifying the musical era, and a direct comparison between the two sets of experiments is difficult.

Introduction

It is obvious that there are overlapping characteristics among different musical eras. Yet, the musical era of a certain piece can be easily recognized by trained musicians. The more one hears compositions of different styles belonging to different eras, the better one can distinguish among them by experience. The human mind extracts certain features in these compositions and categorizes them. This is not to say that the human mind is perfect or infallible, but it degrades gracefully. There might be certain compositions that strongly carry features from two or more eras (e.g., late Beethoven), in which case the trained listener will express doubts about them before coming to a conclusion.

With the developments in the area of connectionism, there has recently been successful use of
Artificial Neural Networks (ANNs) for music related problems like pitch and tonal recognition, and to a lesser extent, music composition (Bharucha & Todd 1989; Scarborough et al., 1991; Todd 1989; Toiviainen 1995). With the correct representation of data, there should be no reason ANNs could not also be applied to higher level cognitive music problems like the recognition of musical style.

The computer-related work so far done in era classification has been non-connectionist and limited to the comparison of the output of two different composers based on musical motifs. Smaill and Westhead (1993) developed a system for automatically characterizing musical style that relied solely on motifs, which they described as patterns of rhythms or pitches common to more than one piece of music in a style. The conclusion from the work was that the motifs alone did indeed provide a way of successfully discriminating between styles as long as the styles were sufficiently different, and the number of compositions analyzed by the system (dictionary size) was large enough.

The goal of the research effort described herein was to develop a method to identify the musical era of compositions using ANNs on a computer, and compare the results to a similar classification experiment on human subjects not having a musical background. The specific objectives of the research for this paper were:

a) To develop a method of extracting the characteristics of a given musical piece, starting from sheet music and ending in digital data to be input to an ANN;
b) To determine the optimal ANN architecture and configuration suitable for training in order to recognize pieces from different musical eras, and train it to classify musical pieces;
c) Perform a similar listening experiment on human subjects not proficient in music and compare their performance to that of the computer.

The Computer Implementation: Model Development

One initial restriction employed before a realistic implementation of an ANN could be developed was that only keyboard music was considered. This reduced the composition set that can be processed to those having only two staves, while still enabling a wide variety of choices from the three eras during which pieces were composed for the harpsichord and the piano. This restriction also eliminated the consideration of the Medieval and Renaissance eras, from which very little keyboard work exists. The Impressionist and Modern eras were also eliminated on the grounds of vastly increasing musical complexity and the broadness of differing musical styles, which would seriously complicate the choice of compositions and the training of the ANN.

Up to the first two pages of each piece were entered into the computer using a scanner and converted to MIDI format using the PIANOSCAN sheet music recognition software. Certain recognition errors were manually corrected and the resulting MIDI file was then converted to a text file that included information about each note of octave and duration. This made it possible to parse the pieces using custom written software and extract the information in any desired way. Due to the limitations of the PIANOSCAN software, any grace notes and loudness information were omitted.
Table 1. Selection of compositions from the three musical eras

<table>
<thead>
<tr>
<th>Baroque Data Set</th>
<th>Classical Data Set</th>
<th>Romantic Data Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>J.S. Bach (Inventions)</td>
<td>Haydn (Selection of Sonatas)</td>
<td>F. Chopin (Selection of Mazurkas, Nocturnes, Preludes, and Waltzes)</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>G.F. Handel (various harpsichord pieces)</td>
<td>Mozart (Selection of early Sonatas and Rondos)</td>
<td>F. Liszt</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>Beethoven (Selection of early Sonatas and Sonatinas)</td>
<td>14</td>
<td>R. Schumann</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td></td>
</tr>
</tbody>
</table>

Although music is a sequential process, a recurrent network was not chosen for this problem due to the difficulties it brought in representing the variable number of notes that can be played at a time (Todd 1989; Atalay 1996). Instead, emphasis was given to the development of an effective feature extraction method to prepare the input data and use a static ANN (Atalay 1996). This method relied on three distinct characteristics of compositions: melody, which is associated with the notes and the order in which they are played; harmony, which is associated with the combinations of notes that are played at the same time; and rhythm, which is associated with the duration values of the notes.

The feature extraction method made use of the following frequency distributions for each hand:

1. **Note Distribution.** All pieces were transposed to the key of C, octave information was discarded, and the number of occurrence of each of the twelve possible notes of the chromatic scale was recorded. The values were normalized between zero and one, so that the most occurring note had the value of one, the rest having values smaller than one in the original ratios.

2. **Duration Distribution.** The same was done for the durations of notes. Twelve possible duration values from the 32nd note to a dotted whole note were allowed (Atalay 1996). All other durations not directly corresponding to the allowed twelve were rounded to the nearest one.

3. **Interval Distribution.** Whenever there occurred more than one note being played at the same time, the number of chromatic intervals was calculated and recorded in the distribution in the same way. For example, the occurrence of the chord C-E-G would give rise to two interval values: four (C-E), and three (E-G). Thirteen intervals (an octave plus half a voice) were permitted as this is usually the limit of what can be played by a single hand.

The loss of octave information in the note distributions was partly recovered by the interval distributions. Although sequential information was not considered, the differences in rhythm in
the two musical eras were captured by the duration distribution to some extent. Also, although a sequence like C,E,G and the simultaneous occurrence of the same notes in chord C-E-G would contribute to the note distribution information in the same way, the difference was captured by the absence of the former in the interval distribution; only the order of the sequence was lost. For example, this made it possible to distinguish between the Alberti bass typical of the Classical era and the left hand chord accompaniment typical of the Romantic era.

Results

Four different comparison cases were tried: Three binary ones, Baroque-Romantic, Baroque-Classical, Classical Romantic, and Baroque-Classical-Romantic. The following table shows the results obtained:

<table>
<thead>
<tr>
<th>Era Recognition Problem</th>
<th>No. of Samples in the Training Set</th>
<th>No. Identified Correctly</th>
<th>Training Set Accuracy</th>
<th>No. of Samples in the Test Set</th>
<th>No. Identified Correctly</th>
<th>Test Set Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baroque-Classical</td>
<td>66</td>
<td>66</td>
<td>100%</td>
<td>24</td>
<td>23</td>
<td>95.8%</td>
</tr>
<tr>
<td>Classical-Romantic</td>
<td>74</td>
<td>72</td>
<td>97.3%</td>
<td>16</td>
<td>14</td>
<td>87.5%</td>
</tr>
<tr>
<td>Baroque-Romantic</td>
<td>46</td>
<td>46</td>
<td>100%</td>
<td>16</td>
<td>15</td>
<td>93.8%</td>
</tr>
<tr>
<td>Baroque-Classical-Romantic</td>
<td>111</td>
<td>105</td>
<td>94.6%</td>
<td>24</td>
<td>20</td>
<td>83.3%</td>
</tr>
</tbody>
</table>

Network evaluation table.

The Human Implementation: Listening Test

In order to put the computer's performance into perspective, it was suggested that some comparison be made with human intelligence as it tackles the subtle task of connecting musical examples with periods in music history. A group of 21 non-music major college students taking a core curriculum Music Appreciation course were asked to choose from Baroque, Classical, or Romantic in response to 18 pieces of music played to them on cassette tape, most of which were the same ones as those used in the computer experiment. The following table shows the list of compositions used:

<table>
<thead>
<tr>
<th>Composer</th>
<th>Composition</th>
<th>Era</th>
<th>% who answered Baroque</th>
<th>% who answered Classical</th>
<th>% who answered Romantic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bach</td>
<td>Invention in C</td>
<td>Baroque</td>
<td>66.67</td>
<td>33.33</td>
<td>0</td>
</tr>
<tr>
<td>Handel</td>
<td>Allegro in A minor</td>
<td>Baroque</td>
<td>28.57</td>
<td>47.62</td>
<td>23.81</td>
</tr>
<tr>
<td>Bach</td>
<td>Invention in F minor</td>
<td>Baroque</td>
<td>23.81</td>
<td>57.14</td>
<td>19.05</td>
</tr>
<tr>
<td>Handel</td>
<td>Capriccio in G</td>
<td>Baroque</td>
<td>38.10</td>
<td>52.38</td>
<td>9.52</td>
</tr>
<tr>
<td>Bach</td>
<td>Invention in A minor</td>
<td>Baroque</td>
<td>19.05</td>
<td>47.62</td>
<td>33.33</td>
</tr>
<tr>
<td>Haydn</td>
<td>Capriccio in G</td>
<td>Classical</td>
<td>19.05</td>
<td>80.95</td>
<td>0</td>
</tr>
<tr>
<td>Haydn</td>
<td>Variations G major</td>
<td>Classical</td>
<td>33.33</td>
<td>57.14</td>
<td>9.52</td>
</tr>
<tr>
<td>Mozart</td>
<td>E♭ Major</td>
<td>Classical</td>
<td>52.38</td>
<td>42.86</td>
<td>4.76</td>
</tr>
</tbody>
</table>
Some students found the task virtually hopeless doing no better than chance in some cases. The best student, however, did almost as well as the computer and on certain items, the students did much better than the computer: The Diabelli Variations of Liszt, for example (See figure 5). The overall accuracy of the students was around 57 percent.

The computer, using technical criteria based on examples, did better than the students with the pieces used in the computer training. The students used in this experiment had no experience (in class) with the specific keyboard pieces selected. They had, nevertheless, a lot of experience with music of many genres in the three periods. So the students relied on how the pieces “felt” (on overall psychological effects such as volume or texture) and made mistakes with pieces such as the Chopin Mazurka in B♭ Major which they characterized as classical and the Bach Inventions (A minor and F minor) which they also characterized as classical. The C major Invention, however, with its more sparse texture was correctly characterized as Baroque.

It would be interesting to test students of various types with pieces of more varied genre and compare their answers with those of the computer when more subtle and/or more complex aspects of music are considered.

There were a number of reasons why the students were at a disadvantage in competing with the performance of the computer: The students had not studied these specific pieces or the test. Instead, they had studied musical styles ranging from Bernart to Bernstein and would be expected to rely heavily on instrumentation to help distinguish musical styles. The computer, of course, concentrated only on music from Bach to Liszt perceived as MIDI input from piano simulations.

For a correct evaluation of this research, it is important to keep in mind its goals. As far as converting sheet music into an ANN input through the MIDI format is concerned, that was done with success within the limitations of the sheet music recognition technology and the MIDI format itself. To evaluate the success of the musical feature extraction method developed and the
ANN architecture used, one has to look at the results of the Backpropagation network on the training and the test sets.

Unfortunately, it would be difficult to compare these results to the performance that could be shown by trained musicians in real life, and the existence of the results of any similar experiments performed on humans would be immaterial. An identical experiment replacing the ANN with a trained musician necessarily cannot be performed, because it is impossible to come up with a human subject who has only heard the compositions in the training sets that were used in this research, and nothing else. Every trained musician must have listened to a greater number of compositions from a greater number of different composers from each era than the ones used in this experiment. Furthermore, they must have heard those compositions in different ways. They might have heard some of them in their entirety, not just the first two pages of the scores. They may have heard the others in a more random way, possibly from the beginning to somewhere in the middle, or from somewhere in the middle to the end.

Also, humans receive music by hearing, which is a distinctly different form of input than the one that was used in this research. Hearing involves a set of very complex processes consisting of numerous physical and biological phenomena, and the perception system for sound information in humans is much more sophisticated than a simple Backpropagation neural network. Still, the good accuracy levels achieved with the simple model for our restricted case implies that musical era recognition might not require a highly sophisticated approach.

References


Computational Auditory Scene Analysis as a Modeling Framework in Musicological Research on Perceptual Auditory Organization

Lüder Schmidt, Uwe Seifert & Randolph G. Eichert
Institute of Musicology, University of Hamburg, Germany

Introduction
Cognitive Science tries to develop a general theory of artificial and natural cognitive systems. Therefore it studies perceptual and cognitive systems in different domains. The main assumption of cognitive science is that the cognitive and perceptual processes are physically embodied in natural systems through computations carried out by the nervous system. Hence, it tries to develop a computational theory of "mind" [18]. The research strategy of cognitive science relies heavily on computer simulation of the assumed computational cognitive and perceptual processes ([34], pp. 307–311). As pointed out by Leman ([22], ch. 13) and Seifert [32], the study of music theory has come to involve aspects of cognitive science, including computer modeling.

In recent systematic musicology, we may roughly distinguish three types of approach towards a description of principles underlying music. The first of these broad types of music theory deals with the problems of internal or mental representations of musical structures, comprising such diverse areas as the linguistically inspired rule system put forth in "A Generative Theory of Tonal Music" [24], attempts to explain tonality and the ordering of musical pitch by internal templates [19] or elaborate geometrical constructions interpreted as "mental representations" [33]. A related approach may be seen in the consideration of these phenomena in the frame of "schema theory" [22].

Even an attempt to understand the historical development of the form of cadences in terms of learning and gradually relearning of auditory patterns (change of templates / schemata / internal representations?) [12] should be regarded as adaptation of the "cognitive" approach.

The first type music theoretical research is contrasted and supplemented by a second one that is more psychophysically oriented. Even in this body of research, the problems of pitch and pitch organization are addressed, but from a different point of view. By way of examples we may mention the theories of virtual pitch (e. g. [38, 39]) or missing fundamental, sensory aspects of consonance and dissonance (e. g. [35, 1]), different attempts to determine the essential features of timbre (review: Handel 1995). Tonality was addressed via the explanation of harmony by psychoacoustic determination of chord roots [36, 31].

In our context, the third, purely mathematical way of describing musical structure (Mazzola and his school) can be disregarded, as it is explicitly not primarily concerned with human perception and cognition [26, ch. 3]. Both the "cognitive" and the "psychophysical" approach have been severely criticized. The psychophysical view has been accused of being reductionist and atomistic, dealing mainly with stimuli that are not appealing to musically educated listener. It is said to neglect the mental activity involved in listening to music as well as effects of context. Partly, psychoacoustical findings are considered as irrelevant for music theory, e. g. the phenomenon of sensory dissonance for the musical problem of consonance / dissonance (cf. the abounding literature, e. g. [9]). This view on psychoacoustics may probably – at least in part – be attributed to the predominance of the "scale oriented view" of psychophysics as opposed to psychophysics as the study of the achievements of our sensory systems (cf. [25]).

The "cognitive" approach on the other hand has been criticized for being too far removed from the physical world – the acoustic input data – by working mainly on the "musical surface" [18]. In the words of Leman ([21], p. 8): "In the past, models of music perception and cognition have too often been based on a skeleton of symbols. It is my personal opinion that music cannot be understood without taking into account its acoustic foundation and auditory (pre)processor: the ear."

Leman [23] therefore advocates what he calls the Convergence Paradigm, assuming "that the correlative study of behavioral data, physiological data and data from computer simulation
may contribute to a better understanding of the processes that underlie music perception and
cognition" (ibid., p. 1).

According to our intuition – at the moment we cannot really call it anything else – what is
now called computational auditory scene analysis (CASA) may yield a framework that allows for
a more precise formulation of the problems aimed at by Leman, giving some hints for solutions,
even if a tremendous amount of work remains to be done. In the following, we'll present by way
of examples some phenomena of musical relevance discussed in the context of auditory scene
analysis (ASA). Problems remaining unsolved will show the need for further clarification and
formalization – again from the musical point of view. This clarification can be regarded as a step
towards computational theory of audition (cf. [17]). But first we'll briefly touch upon some
concepts of ASA.

Auditory scene analysis
The term "auditory scene analysis" (ASA) was introduced by Bregman [5, 6] for studies
concerning the problem of how an auditory system extracts information from the physical signal
to create a description of the world surrounding the subject. In related work (e. g. by Yost,
Hartmann, and Moore), different terms were used to denote essentially the same area of research,
namely "perception of auditory entities", "auditory object perception", "auditory image analysis",
and "sound source determination" (cf. [41]). To achieve the goals indicated by these expressions,
the auditory system is assumed to decompose the incoming sound field into the constituent
components, which are subsequently grouped according to rules specified in ASA research.

A fundamental concept of ASA is the (auditory) stream, which is defined by McAdams
and Bregman ([27], p. 659) as a mental representation of an acoustic phenomenon: "A stream is a
psychological organization that mentally represents such a sequence [i. e. some sequence of
acoustic events emanating from one location of a physical source] and displays a certain internal
consistency, or continuity, that allows that sequence to be interpreted as a whole."

The formation of streams is the product of two types of grouping. The first, "sequential
integration", addresses the binding together of consecutive (groups of) partials. Cues for
coherence of such consecutive events may be frequency proximity, rate of tone repetition,
similarity in timbre and loudness ... . "Simultaneous integration" denotes the grouping of
simultaneously sounding partials (as in complex tones), fusion being promoted by cues such as
common onset or common frequency modulation. Another distinction is drawn between
"primitive" and "schema-based" stream forming (streaming) processes (segmentation; cf. [5],
especially ch. 4). Primitive stream segregation is taken to be based on innate and therefore
culturally independent principles, whereas schema-based segregation involves prior learning and
the direction of attention. Alternative organizations of the auditory input can arise from
competing principles of stream segregation.

Musical relevance of auditory scene analysis
To give an idea of the interplay of scene analysis rules and the perception and cognition of music,
we will take up a discussion on "Cognitive aspects of amadinda xylophone music" (Wegner [40]).
For additional elucidation we will point out a potential analogy in western music. Further aspects
on western music are discussed in [4; 27; 5, ch. 5].

Nearly 40 years ago, Kubik ([1960] et passim) introduced the notion of "inherent pattern"
to describe a phenomenon experienced when listening to music played by Buganda musicians on
the amadinda. The music consists of two or three parts played on one xylophone. Two players
perform different isochronous repetitive patterns, the note of one occurring half time in between
two notes of the other. (The third player just doubles some of the notes of the other two players.)
When the two parts are joined at the usually rather high tempo, it was observed that they will not
be heard separately but will be perceived as forming new patterns that group together notes from
the parts of both players. The situation may be compared with the two interleaving violin parts at
the beginning of the final movement of Tchaikovsky's 6th symphony: the listener usually does
not perceive the erratic movement of the isolated violin parts but a rather smooth melody and will probably be surprised when casting a look at the score for the first time. Both effects may be explained by the same principles of primitive ASA: as there are no differences in timbre, location, and loudness, auditory streams are formed on the basis of frequency proximity, comprising the melody in the case of Tchaikovsky and the inherent pattern in the amadinda case.

As we would expect from an experienced violinist who has performed the Tchaikovsky symphony a couple of times to be able to follow his own part in an auditory presentation, Wegner [40] found that amadinda players could indeed follow their parts in the presentation of a well known amadinda piece ("Ssematimba ne kikwabanga"). Although this should not come as a surprise, we have here a clear indication that learned patterns / schemata can interfere with primitive stream segregation. Regrettably, Wegner did not set out to specify the exact conditions of his performance regarding frequencies, tempo,... . Thus, a detailed analysis of the relation of schema-based and primitive stream segregation in this context is still lacking. Other findings reported by Wegner [40] point in a similar direction.

Computational auditory scene analysis

Work on computational auditory scene analysis in music was initiated by attempts to devise systems for automatic transcription of acoustic input and "intelligent" editing of recorded material. Starting out with ad hoc engineering solutions, later the need was felt to incorporate the study of the functions of the human auditory system, physiologically as well as psychologically [10, 30]. Another point of departure was work on automatic speech recognition, researchers turning to music in the hope of finding a means for easier control of their implementations by producing symbolic output in form of musical notation, thus being dispensed from the need to perform formal listening / understanding tests ([8]. p. 108).

Marc Leman [21] pointed out the importance of auditory models for music research. Although actually he is not directly involved in research on auditory scene analysis, his research [22] on tone center perception is based on auditory models.

Auditory models are built on data collected from neurophysiological and psychoacoustical studies to extract task-related information. The task in computational auditory scene analysis is to decompose a mixture of sound into elements and then to assemble the elements into single percepts according to some relevant cues. In general, four stages can be functionally differentiated (c. f. Ellis [14], pp. 26 and pp. 66): front-end, basic representation, grouping algorithm and output. The front-end converts the acoustic signal into suitable representations like an onset map or a time–frequency intensity envelope to enable further processing. At this stage of processing detailed knowledge of the physiology of the processing in the auditory periphery like cochlea filtering, the mechanical to electrical transduction exhibited by the hair cells and the neural activity of the auditory nerve fibers is used. Onset detection takes place at this stage. Based on acoustic cues object formation takes place and discrete sound entities are built. At the third stage grouping of the sound entities is carried out. The last stage depends on the specific goal of the system. One example may be output of the sound signal in musical notation.

Although the front-end of the systems often are based on physiological dat of the auditory periphery not all systems are modeled on these data. The scene analysis systems presented so far that have been applied to musical stimulus material, differ in the extent to which physiological and psychoacoustical data has been incorporated.

For example, Cooke [11, 8] presents a neurophysiologically inspired modeling approach to auditory scene analysis. He discusses peripheral and central auditory information processing of the nervous system. This model starts with a modeling of cochlea filtering and the hair cell transduction process. Brown and Cooke [8] used, as Uwe Baumann [2] did in his study, two measures (measures 19–20) of J. S. Bach's "Minuet in B flat" from "A little note book for Anna Magdalenaa Bach" to demonstrate the capabilities of their model to segregate polyphonic music. The separation process is based on timbre. Timbre is represented as a two dimensional space of "brightness" and onset asynchrony.
The approach to auditory scene analysis by Mellinger and Mont-Reynaud [29, 28] tries to combine neurophysiological, psychoacoustical and psychological data. They [29, p. 30; 28, p. 282] use simultaneous saxophone and drum tones to be separated. So in general they avoid the problem of overlapping harmonics which often occurs in polyphonic music ([8], p. 125).

Baumann [2] puts his computational modeling approach to auditory stream segregation into relation to the general framework of hierarchical auditory information processing as developed by Ernst Terhardt [37]. This framework is based on ideas from biological evolution, and Gestalt theory and ethology. Learning and conditional decision making are essential to this framework in order to understand auditory information processing in the nervous system. Baumann's model uses an "aurally adequate" algorithm for spectrum analysis. Evaluation functions based on psychoacoustical principles are the second main component in Baumann's approach. Gestalt principles like common fate, closure, proximity, similarity and good continuation are used as heuristics. First, "acoustical objects" are formed, then "auditory object patterns", "pitch-evaluated auditory patterns" are built and at last "auditory objects are sequentially connected" to represent the melody or voice experienced by a listener focusing his attention to this voice in polyphonic music. Baumann [2, pp. 115-116] uses short extracts from simple polyphonic music like J. S. Bach's "Minuet in B flat" from "A little note book for Anna Magdalena Bach" to show the capabilities and limits of his system to segregate the voices. One of the main problems in stream segregation in polyphonic music is to discriminate simultaneous spectral components of different sounds: the harmonicity problem. Despite psychological consideration Baumann's approach is in essence a psychoacoustical approach to auditory stream segregation. Dan Ellis [14] summarizes previous studies on computational auditory scene analysis and is skeptical about the application of current models of scene analysis to transcribe music automatically [14, pp. 21-22]. He too is skeptical about approaches that are based on neurophysiological data to support auditory scene analysis and advocates a modeling approach based on psychological data as heuristics for implementation and a top down approach in an distributed symbolic black board architecture.

Although Beauvois and Meddis [3] did not use musical stimuli in their computational model on auditory scene analysis, their work seems to be interesting for researchers concerned with musical stream segregation. In music research as in research of auditory streaming it is often assumed that the segregation and fusion processes can be only explained by Gestalt principles. But Beauvois and Meddis showed that this is not always necessary and an explanation based on peripheral physiological data can account for some auditory scene analysis phenomena. They think of their results as being complementary to the high-level Gestalt explanations. From this brief and superficial review of current research on computational auditory scene analysis it can be concluded that despite the fact that some considerable progress has been made in the segregation of (musical) sound by computer and modeling of hearing much remains to be done. Only short parts of "real" polyphonic musical example have been fairly well segregated. Regarding the complexity of the task one should bear further in mind that the main goal of most of these systems has not been to built a model or theory of music perception. Music has been a special test case for evaluating the capabilities of these systems to separate and identify sounds.

Conclusions

In the introduction we mentioned two main positions in research on music: the cognitive and psychophysical approach. The computational approach to auditory scene analysis allows to bring together the sensory, perceptual and cognitive perspectives in research on music perception and cognition. The possible symbolic (mental) representations of musical structure and their processing are met by constraints due to the properties of the physical environment, i.e. in the case of music perception and cognition the neurophysiological realizations of the auditory functions.

Computational auditory scene analysis opens the possibility to develop a structural and processual theory of music cognition. Further, we think that in the realm of computational auditory scene analysis psychological and neurophysiological research can be brought together...
with research on music theory. Apart from this broad theoretical perspectivesome "side effects" may be noted. The computational approach to audition augments the need for a specification and formalization of the concepts of schemata and their action / interaction with other processes in hearing. Not mentioned here (as not explicitly addressed in the models discussed) are the influence of expectations or attention, as e. g. in the context of (musical) rhythm. One popular theme in cognitive research on music has been the application of Gestalt principles. Although they can be accepted as plausible and suggestive descriptions of phenomena experienced, their explanatory value has to be questioned [13]. As indicated by the results of Beauvois and Meddis [3], within a computational approach towards auditory music (perception and) cognition there may gradually emerge strategies of explanation embedding Gestalt principles in a theoretical framework. In our paper we focussed on auditory scene analysis as a means for music research. The importance of research on auditory stream segregation in connection with tonality perception for empirical music theoretical research that is not constrained to a specific culture, epoche or style has been underlined by Erickson [16]. Auditory scene analysis addresses the question how the auditory system builds up a mental representation of a "sound structure" out of a mixture of sounds first by decomposing the sound and building then auditory entities through competitive processes of sequential and simultaneous grouping. Experimental research is complemented by computer simulation and model building to achieve an explicit theory of the principles of auditory scene analysis. Computer simulation of auditory functions allows to relate music theoretical, psychological, neurophysiological and psychoacoustical research. In his article [15] Ellis discussed several problems in computational auditory scene analysis and proposed to use some standard sounds to compare different systems. To us "Ssematimba ne Kikwabanga" seems an interesting starting point to further explore auditory scene analysis in music from different perspectives like music theory, music perception and cognition. For computational approaches to music perception it is a challenge and test case. Several problems could be studied in a computational framework on music perception like scales and categorical perception, pitch recognition of inharmonic sounds, the influence of "consonance" and timbre on melodic relations and stream segregation, the influence of cultural knowledge. Research on computational auditory scene analysis should integrate computational approaches to pitch perception, because principles of pitch perception seem to be strongly related to the perception the chord root, which seems to be central to concepts like harmony, tone center perception and tonality.

References
The Role of Similarity in Categorisation: 
Music as a Case Study 

Emilios Cambouropoulos 
Faculty of Music and Dept. of AI, University of Edinburgh 
emilios@music.ed.ac.uk

Introduction

A commonly encountered hypothesis on which many categorisation models are grounded is that categorisation is strongly associated to the notion of similarity, i.e. similar entities tend to be grouped together into categories.

There are though different views on the relation between similarity and categorisation (Goldstone et al. 1994, Medin et al. 1993). On one hand, similarity is considered to be too flexible and unwieldy to form a basis for categorisation, e.g. any two entities may be viewed as being similar in some respect (e.g. a car and a canary are similar in that both weigh less than 10 tons, but these objects are not normally considered to be members of the same category!). On the other hand, similarity is regarded to be too narrow and restricting to account for the variety of human categories (e.g. a whale is more similar to other fish but we still consider it to be a mammal). Goodman (1972) doesn't hesitate to call similarity 'a pretender, an impostor, a quack' (p.437).

The above debate is directly linked to a further issue, that is how entities and their properties are represented; if objects are described in terms of mainly perceptual (e.g. visual or auditory) properties, then, obviously similarity is insufficient for categorisation, whereas, if any sort of properties - perceptual/abstract/relational - are considered then similarity becomes too flexible.

It seems that the notions of categorisation, similarity and the representation of entities/properties are strongly inter-related. It is not simply the case that one starts with an accurate description of entities/properties, then finds pairwise similarities between them and, finally, groups the most similar ones together into categories. It seems more plausible that as humans organise their knowledge of the world, they alter their representations of entities concurrently with emerging categorisations and similarity measurements. Of course, there are some general perceptual constraints as to what is perceivable in the first place, but from there on different properties of entities become more prominent in a given context for a specific categorisation task or for a similarity judgement (cf. diagnosticity principle in Tversky, 1977).

It is suggested that perhaps much of the controversy on the role similarity plays in categorisation is due to the lack of an agreed understanding of these notions. So, in the first part of this paper some commonly used - usually implicitly - hypotheses/definitions will be outlined and discussed. In the second part of the paper, a working formal definition of similarity and category will be given along with a general dynamically evolving algorithm (unsupervised symbolic machine learning) that classifies entities. As categorisations are refined so are similarities between entities and the prominence of different properties. Finally, some psychological experiments that seem to suggest an incongruity between the notions of similarity and categorisation will be re-visited and alternative interpretations that are compatible to the proposed formal definitions will be given.

Musical Similarity and Categorisation

But how does music relate to this debate on the role of similarity in categorisation? Musical 'meaning' is established to a great extent through self-reference i.e. the relations of new musical passages to previously heard material. Similar musical elements are organised into musical 'categories' such as rhythmic and melodic motives, phrases, themes and variations, harmonic progression groups etc. But when are two different musical passages similar? And when are two passages different enough as to be considered dissimilar (or contrasting)? Which musical passages belong to the same paradigm/category? What happens with ambiguous passages?

One of the main assumptions made in this paper is that similarity always depends on context (when similarity seems to be relatively stable, this is so simply because the context -e.g. the
structure of the natural world—tends to be quite stable). A musical work may be considered as a local context within which things like motives, themes, harmonic progression groups etc. emerge. Trying to discover the similarity of two isolated musical passages will usually produce dubious or relatively uninteresting results. Consider, for instance, the musical passages below. In which of the two pairs the two passages are more similar?

Some might select the first pair, others the second pair, and still another group might refuse to make a judgement. It is suggested that perhaps this similarity experiment is simply ill-designed in the first place, and perhaps subjects of the third group are right in refusing to make a judgement. The problem seems to be that these excerpts are taken out of their context. As it happens, the first two passages are very dissimilar—actually contrasting—within the homogeneous minimal context of S. Reich's *Electric Counterpoint*, whereas the second two are very similar within the very diverse context of I.Xenakis' *Keren*. Context seems to be paramount in our establishing similarities and categories between musical passages and it is asserted that it is not possible to find an absolute criterion for defining what things are similar in general.

**Entities/Properties, Identity, Similarity and Categorisation**

An *entity* is taken to refer, in this text, to a complete and distinct thing—concrete or abstract—such as an object, a structure, a function, a goal, and so on (e.g. a note, a song, a dominant chord, an emotion, etc.). An *property* is any predicate that may be used to describe an entity.

Similarity is very often defined as *partial identity* i.e. two entities are similar if they have some predicates the same but not necessarily all. This follows from the definition of identity: for a given domain of discourse, two entities are identical if the same predicates (properties) are satisfied by both (Russell, 1964, vol. i, def. 13.01; see also Quine, 1950). Similarity between two entities may be calculated by simply counting the number of matches between their properties. In general, similarity may be defined as a multivariate function of the pairwise differences between all the properties (weighted or not) these objects posses; usually similarity is inversely related to the distance between the two entities (Polansky, 1996). There exist also asymmetric definitions of similarity (Tversky, 1977; Krumhansl, 1978).

In the course of this text the word *category* will be taken to refer to a set of entities which are grouped together on the basis of some criteria/rules. According to the classical *monothetic* definition a category is constituted of all the entities that posses a set of properties or satisfy a set of conditions (see Sutcliffe, 1993). Most commonly these conditions are taken to be singly necessary and jointly sufficient.

A different approach to formalising the notion of categories has emerged following Wittgenstein's approach to the notion of 'family' and 'family resemblance' (Wittgenstein, 1953). According to the 'modern' *polythetic* view, a category consists of individuals that have a large number of properties from a given set \( P \) and that each property is possessed by a large number of members but no property is possessed by all the members of the category (see Beckner, 1959, p.21). The polythetic definition of categories underlies prototype models of categorisation (Rosch, 1978; Hampton, 1993), exemplar models (Estes, 1994), indirect clustering models (Murtagh, 1993) etc. This definition seems to require a threshold/limit that determines the lowest number of properties that should be shared by each pair of category members i.e. one should be able to determine when a 'large number' is large enough.

It is often stated that models based on the polythetic definition of category reflect human categories in a better way because boundaries of categories may be unclear or fuzzy (e.g. is a tomato a fruit or a vegetable?). The polythetic definition, though, allows sharp boundaries if an exact threshold is defined. It is suggested herein that ambiguity in human categorisations may well be modelled by using categories with sharp boundaries if overlapping of categories is allowed. The more categories an entity belongs to, the more ambiguous it is.
It is clear from the above discussion that all the members of a category are necessarily pairwise similar as they necessarily share some common properties, but the reverse is not necessarily true, i.e. similar entities are not necessarily members of the same category. The notions of similarity and category can be brought into a 'one-to-one' relation if a threshold is introduced in the definition of similarity (see below).

In the following section a formal working definition of similarity and categorisation will be presented along with a dynamic process for determining similarities and categories when an initial set of entities and properties is given.

**Working definitions and a computational model for categorisation**

Let \( T \) be a set of entities and \( P \) the general set of all the properties that are pertinent for the description of each entity. If \( d(x,y) \) is a distance/proximity metric with which the distance/degree of difference between two entities may be calculated, then similarity between them is defined as:

\[
  s_h(x,y) = \begin{cases} 
  f_h(d(x,y)) \geq 0 & \text{iff } d(x,y) \leq h \\
  f_h(d(x,y)) < 0 & \text{iff } d(x,y) > h 
  \end{cases}
\]

where \( s_h(x,y) \geq 0 \) means that the two objects are similar for a given \( h \) and dissimilar for \( s_h(x,y) < 0 \), \( h \) is a distance threshold \((0 \leq h \leq d_{\text{max}})\) and \( f_h(d(x,y)) \) is a monotonic decreasing function of distance such as \( f_h(d(x,y)) = h - d(x,y) \). (Similarity may also be: \( s_h(x,y) = 1 \) iff \( d(x,y) \leq h \); 0 iff \( d(x,y) > h \))

The above definition of similarity is brought into a 'one-to-one' relation with a notion of category. That is, within a given set of entities \( T \), for a set of properties \( P \) and a distance threshold \( h \), a category \( C_k \) is a maximal set with \( n \) entities:

\[
  C_k = \{ x_1, x_2, \ldots, x_n \} \text{ such that: } \forall i,j \in \{1,2,\ldots,n\}, s_h(x_i,x_j) \geq 0
\]

In other words, a category \( C_k \) consists of a maximal set of entities that are pairwise similar to each other for a given threshold \( h \). A category, thus, is inextricably bound to the notion of similarity; all the members of a category are necessarily similar and similar entities define a category.

The above definition of category and similarity readily lends itself to form the basis of a dynamic process for discovering pertinent categories and similarities \((0 < h < d_{\text{max}})\), when accompanied by the general principles of economy and informativeness (Cambouropoulos 1997).

When a threshold is chosen, then the initial weights of properties can be altered so as to optimise the distinctiveness of the category's description. Weights for each property \( A_{ij} \) (\( i \)th value of attribute \( A_i \)) may be adjusted in relation to the diagnosticity of that property for a given category i.e. properties that are unique to members of one category are given higher weights whereas properties that are shared by members of one category and its complement are attenuated. In other words, the dimensions in a multidimensional space are adjusted in such a way that distances between members of different categories are maximised. Such a function could be:

\[
  w_{ij} = |m/n - m'/N-n| \quad \text{where, } m = \text{number of objects in category } C_k \text{ that posses attribute } A_{ij}, \\
  m' = \text{number of objects not in category } C_k \text{ that posses attribute } A_{ij}, n = \text{number of objects in } C_k, \\
  N = \text{number of objects in } T.
\]

The weights of each property calculated for each category can then be averaged and normalised for a given categorisation description. If an acceptable classification has not been arrived at, the whole process may be repeated for the new set of weighted properties until a satisfactory categorisation is achieved. The selection criteria for the best categorisation descriptions given for the various thresholds \( h \) may be among others: a) an exhaustive description of the object set, b) minimum overlapping between the categories, and c) the lowest and highest values of \( h \) are avoided being too specialised (each object a category of itself) or too general (all objects form one category) respectively.

**A Musical example**

Paradigmatic analysis (Nattiez, 1975) is concerned with the organisation of a monophonic piece...
into vertical columns of similar musical segments. The following musical segments appear in
Nattiez's paradigmatic analysis of Debussy's *Syrinx*:

```
A B C D E F G
```

Segment D is placed by Nattiez in the column with motives E, F and G although one might
initially think it would be more obvious to place segment D with A, B and C. How would this
limited set of musical entities be categorised according to the above similarity/categorisation
algorithm?

Let's assume we have a rudimentary set of pitch-interval and duration parametric profiles for
each of these musical segments i.e. exact pitch intervals (in semitones), contour and durations:

```
Am: {rh1, rh2}  A_{pex}: {pex1, pex2, pex3, pex4}  A_{pcont}: {pcont1, pcont2}
```

If the initial weights for all the properties are $w_{ij}=1$, we have the following categories (similarity
values are not depicted) according to the similarity/categorisation algorithm (there are 4 possible
distances therefore 4 useful thresholds):

- **Threshold: $h=3$** → Categories: $\{A,B,C,D,E,F,G\}$
- **Threshold: $h=2$** → Categories: $\{A,B,C,D\}, \{D,E,F,G\}$
- **Threshold: $h=1$** → Categories: $\{A,B,C\}, \{D,E\}, \{E,F,G\}$
- **Threshold: $h=0$** → Categories: $\{A,B,C\}$

If some overlapping is allowed then the two descriptions for $h=2$ and $h=1$ are acceptable
according to the selection criteria. The description for $h=2$ is somewhat simpler so preferable. It
is obvious that segment D is ambiguous as it can be placed with $\{A,B,C\}$ and/or $\{E,F,G\}$.

If no overlapping is allowed then one might select the most stable category $\{A,B,C\}$ for $h=0$,
calculate new weights for the attribute set ($w_{rh1}=0.75$, $w_{rh2}=0.75$, $w_{pex1}=1$, $w_{pex2}=0.5$,
$w_{pex3}=0.25$, $w_{pex4}=0.25$, $w_{pcont1}=1$, $w_{pcont2}=1$) and then apply the similarity/categorisation
algorithm to the segments for the new weights. This yields among other classifications:

- **Threshold: $h=0.68$** → Categories: $\{A,B,C\}, \{D,E,F,G\}$

This conforms with Nattiez's preference in placing musical segment D with the segments of the
column/category that includes segments E, F and G. From the above weights it is clear that, for
this classification, contour and pitch pattern pex1 are more diagnostic.

The process could have started with different initial attribute weights, e.g. the attribute 'rhythm'
could have double weight (this would be quite reasonable in the sense that rhythm and pitch
profiles would be overall equally important). In this case among other classifications we have:

- **Threshold: $h=5$** → Categories: $\{A,B,C,D\}, \{D,E,F,G\}$
- **Threshold: $h=2$** → Categories: $\{A,B,C,D\}, \{E,F,G\}$

In this case, where the initial weight of the attribute 'rhythm' is higher, the musical segment D is
categorised with segments A, B and C (for $h=2$), if no overlapping is allowed, as one might
have initially guessed (the attribute weights in this case are: $w_{rh1}=1$, $w_{rh2}=1$, $w_{pex1}=0.75$,
$w_{pex2}=0.08$, $w_{pex3}=0.33$, $w_{pex4}=0.33$, $w_{pcont1}=0.75$, $w_{pcont2}=0.75$).

The set of weighted attributes for each category along with the range of thresholds for which this
category occurs can be used to make membership predictions of new unseen musical segments
Cambouropoulos, 1997).

This musical example illustrates the flexibility and adaptiveness of the proposed similarity/
categorisation algorithm. Segment D can either be grouped with segments $\{A,B,C\}$ or with
segments $\{E,F,G\}$ depending on the initial weighting of the musical parameters or may simply
be considered as an ambiguous hybrid of the two classes (although most analytic theories that are
based on strict hierarchic non-overlapping descriptions would reject ambiguous overlapping
descriptions). When human analysts make a paradigmatic analysis of the same musical piece it is
almost certain that they will arrive at different descriptions. This is due to the fact that each
analyst gives different prominence to the various musical parameters or might even use
somewhat different parameters altogether and, of course, may choose different thresholds for what is considered to be similar/dissimilar. All of these possibilities are accommodated in the proposed system of categorisation.

Re-examining some psychological experiments

We will now examine how the notions of identity, similarity and categories have been applied in three psychological studies/experiments and will shown that these experiments need not be considered incompatible with the proposed working definitions.

1. Krumhansl (1990, pp. 143-152) suggests that two instances of the same musical tones are perceived as being more self-similar or identical if they are more stable in a given tonal context. The first principle, contextual identity, governs the degree to which two instances of a musical tone are perceived as identical... For two instances of the same tone, a, the psychological distance is denoted d(a,a). The principle says that this distance is less for more stable tones. Contextual identity: d(a,a) decreases as the stability of 'a' increases' (Krumhansl, 1990, p. 143).

One experiment conducted by Krumhansl that supports the above principle, involves listeners comparing and measuring the degree of sameness/difference of the two instances of the same tone preceding and following the same tonal context For example, a middle G is played before and after a C major context and a middle F# before and after the same C major context. Listeners gave a higher rating of sameness to the more stable diatonic tone G than to the non-diatonic tone F#. Although both instances of the two tones have the same pitch and both occur in the same tonal context (i.e. they are identical) they are judged to be to a different degree identical.

According to the definition of identity given previously, two entities are identical if they share all the same predicates in a given domain of discourse. This means that two entities in a given context are either identical or not - there can be no degree of identity. In the light of this definition the use of the term identity in relation to the above experiment is questioned. It is suggested that the two instances of a tone presented to listeners in the above experiment are in the first place non-identical as they occur in different temporal positions in relation to the given C major tonal context. Actually the first occurrence of the tones doesn't have any local context except in retrospect: it may be hypothesised that the listener makes a tentative assumption according to background knowledge that the first standard tone is a tonic or another diatonic tone which may be overturned by the subsequent context - contrastingly, the last tone is clearly placed in relation to the preceding context. Perhaps one way to have the 'same' context for both instances of the tone is to present listeners the sequence: Cmaj context - tone X - Cmaj context - tone X (possibly looped indefinitely).

Perhaps this issue could be resolved if the word identity appeared in inverted commas, as the principle of contextual 'identity'.

2. Carrey (1985) presented to subjects a set of living things that include one mechanical monkey. Then subjects were asked to select an item from this set that was most similar to a human; both children and adults chose the mechanical toy monkey. However, when they were asked about the biological properties of the mechanical monkey all subjects denied that the mechanical monkey had any at all (e.g. it doesn't have a heart, it doesn't sleep etc.). So, although the mechanical monkey was judged to be most similar to humans the two were not considered to be members of the same category. Murphy (1993) refers to this experiment as an 'impressive demonstration' that it is not generally the case that 'the more similar an object is to a conceptual representation, the more likely it is that it will be identified as an exemplar of that concept' (Murphy, 1993, p.185).

This experiment - if adequately interpreted - seems to be in line with the claims of this paper on the strong link between similarity and categorisation. One interpretation of this experiment is that perceptual visual similarity (appearance) is not generally sufficient (or even relevant) for categorisation. But it doesn't seem to support Murphy's claim. Visual similarity is not the same thing to more general conceptual similarity. Another way to view this experiment is that subjects are not making judgements of perceptual similarity, but are simply using the mechanical toy monkey as a signifier/sign for a real monkey - since all the other objects are living things - and are actually comparing a real monkey to a real human. In this case, subjects may be making use of a more general notion of similarity and actually the experiment may be taken to be in support
of the claim that similarity is strongly bound to a notion of categorisation. (Perhaps, if the order of the experimental stages was reversed, i.e. first the discussion on biological properties and then similarity ratings, then the actual mechanical monkey might have not been judged to be similar to the human, as subjects would probably be using a broader notion of similarity.)

3. Barsalou's (1983) 'ad hoc' categories are often used as examples of categories whose members are dissimilar. Examples of such categories are: 'things to take on a camping trip', 'foods to eat on a diet', 'things to take from a burning house' and so on. It is suggested that such categories rely on transient goals rather than similarity between the objects. But such goals may be considered as properties of the objects in a given domain of discourse in which case the objects are similar as to these goal-oriented properties. Murphy states that "Children, jewellery, family photographs, and pets are quite different in most respects, but they are similar in that they are portable, people value them highly and they are irreplaceable. Thus, they are all excellent candidates for 'things to carry out of a burning house'." (Murphy, 1993, p.186)

Conclusion

In this paper, a working formal definition was given according to which similarity is always relative/context-dependent and is inextricably bound to a notion of corresponding categories. This definition was used as the basis for a dynamic process whereby, given a set of objects and properties, a range of plausible classifications of similar entities for a given context is generated and the most diagnostic properties highlighted. This model allows membership ambiguity in terms of overlapping of categories, automatic determination of number of categories, explicit description of categories and membership prediction of new entities. A simple musical example was presented that illustrates the capabilities of the model. Finally, some examples/experiments from psychological research that seem to contradict the proposed definitions were critically re-examined and it was shown that they are compatible with the current description of similarity and categorisation.

References


Toward a formal model of musical perception

Pierre-Yves Rognin
Laboratoire d'informatique de Marseille
Faculté des sciences de Luminy
163, Avenue de Luminy
13288 Marseille, Cedex 9, France
e-mail: Rognin@univ-avignon.fr

Abstract
This paper is an attempt to formalize fundamental aspects of musical perception. In the long-term, we want to simulate the production by a listener of a musical piece representation. There are two starting points: a J. Kunst article in which a modal logic based model is presented, and a study of the U.S.T. notion (Temporal Semiotic Unity) proposed by the M.I.M. (Music and Computer Science Laboratory of Marseilles).

We propose an implementation of the Kunst model, U.S.T. being considered as the vocabulary underlying the logical language of the model.

1 The Kunst model
1.1 General framework
The formalized musical understanding process concerns knowledge revision. When hearing a musical piece, the listener builds a mental representation of it. But he can change his strategy when he perceives some new information. Moreover, his new interpretation depends on what he heard before. The problem is to obtain a model for the description of the dynamic aspect of memory during the listening phase.

Kunst proposes the bivalence function (BF) to represent this passage of one musical understanding to another (UNLL : UNLearning plus Learning). Logical formulas represent the different interpretations of the listener. These formulas are the arguments of the function.

1.2 Formalism

- Modal logic: The temporal modal logic will be the mathematical framework of the formalisation. This is a logic based on propositional logic, with a modal operator □. □ and its dual ◊ are defined as:

□p : The music has now behaviour p and has always had it in the past.
◊p : The music had behaviour p in one or more possible pasts.

At a given time, the different strategies of the listener (for characterizing music) are represented by possible pasts. If an interpretation p is true in every possible past, we will call p a musical law at this time, and it will be noted □p.

• **Axiomatics**: The modal system used is S7, essentially because of three axioms: reflexivity, transitivity and non-normal worlds axiom. The first axiom means that a listener knows what he is perceiving when he is perceiving it. The second axiom means that the listener’s memory works perfectly. The last axiom means that each musical interpretation comes from an arbitrary decision. From Kripke’s semantic point of view, this last axiom implies some non-normal worlds: any interpretation is possible in them.

• **The bivalence function**: Kripke’s semantic is used to represent a bivalence function (BF). In terms of logic, a BF is an S7 model.

The BF(p,q) is represented by figure 1. It corresponds to a musical situation considered as a problem resolution for the listener. A possible reading is proposed below:

- The starting point in time is w₀₁. At this time, a law is formulated: the music has behaviour p. This law refers to a musical past w₀₀: a music class with behaviour p is chosen as a specific context for the music.

- Later, an element is perceived, which implies p is no more a musical law. This is indicated by w₁₂(1) (wp : world perception). And this is reflected in w₀₂: p was a musical law at one time (□p), but it is no more the case (¬p).

- The listener is at the first stage of UNLL: UNLearning. But in w₁₂ there is some new information: (2) ¬p ∧ q. At the light of this information, the listener builds a new past for the music (w₁₁, w₁₀), and finds a solution for w₀₂: the musical law □q in w₁₂(2). This is the learning part of UNLL.

We can summarize the structure of a BF as follows:

- The boxes represent points in musical space-time, the time flow running from top to bottom.

- A double-lined box is the starting point of an interpretation. This starting point is either situated at the beginning of or during the listening process.

- The whole block diagram (figure 1) is a S7 model from Kripke’s semantic. Boxes represent possible worlds. Double-lined boxes are non-normal worlds, arrows represent accessibility relations between worlds.
2 The U.S.T.

- **Definitions**: U.S.T. occurrence: Musical segment with temporal meaning, following a morphological organisation, not context-dependent.
  
  U.S.T.: equivalence class of musical segments with same temporal meaning.

- A musical organisation of time: U.S.T. are hearing analysis elements, created by the M.I.M. The M.I.M. research is twofold: first it is an attempt to develop some description tools of elementary temporal organisations, particularly for non-measured musics. Secondly, this research extends P. Schaeffer's theory of "sound objects" ([Chion]) in a semantic way. That is to say that these musical objects have an intrinsic meaning, outside any musical context. From different musical pieces, the M.I.M. have isolated some musical creatures, which seemed to have an intrinsic meaning. Then they made different clusters to produce the abstract definition of U.S.T..

- Components: The M.I.M. have proposed a systematic description of each U.S.T. with the help of characteristic components: morphological and semantic ones.

<table>
<thead>
<tr>
<th>Morphological components</th>
<th>Semantic components</th>
</tr>
</thead>
<tbody>
<tr>
<td>duration</td>
<td>delimited or not</td>
</tr>
<tr>
<td>reiteration</td>
<td>orientation</td>
</tr>
<tr>
<td></td>
<td>with or without</td>
</tr>
<tr>
<td></td>
<td>motion</td>
</tr>
<tr>
<td></td>
<td>with or without</td>
</tr>
</tbody>
</table>

*Table 1: Morphological and semantic components (extract)*

3 The model implementation

We carry out a first simulation through U.S.T., with a Haydn sonata as a musical example.

3.1 The automatic part

BFs are S7-logic models. Firstly we present a tool for studying models in this logic, a S7 demonstrator. We will then see how to build BF.

- **A demonstrator for S7**: The demonstrator uses the semantic boards method [Hughes & Cresswell 68] for an explicit construction of S7 models.

  - Principle: Let $L$ be a logical system, $M$ a set of Kripke models, and $\alpha$ a formula of $L$. If $\alpha$ is not a theorem of $L$, $\neg \alpha$ must be satisfiable in an $M$-model. Otherwise, the construction of a model for $\neg \alpha$ must necessarily encounter a contradiction.

  - Structure and principal rules: the representation is inspired from [Catach 91]. A S7 model will be a quadruplet $< T, R, W, Q >$. $W$ is the set of worlds, $R$ a binary relation on $W$, $Q$ the set of non-normal worlds of $W$. $T$ is a set of triplets $< w, f, a >$ where $w$ is a world of $W$, $f$ a formula of $w$, and $a$ is equal to 1 or 0 depending on whether the formula is processed or not. When the algorithm begins, $W = \{ w_0 \}$, $Q = \emptyset$, $R = \emptyset$, $T = \{ < w_0, \neg \alpha, 0 > \}$. We want to know if $\alpha$ is a theorem.

  There are general rules such as:

  (R*1) If a world is reached, in which there is only $\Diamond f$-formula, then this world becomes non-normal.

  (R*2) The first world is normal, with a link to a non-normal world.

  And there are transformation rules such as:

  (R1) $< w, \square f, 0 > \rightarrow < w, \square f, 1 >$, $\{ < v, f, 0 > / (w, v) \in R \}$

2Laboratoire Musique et Informatique de Marseille
Algorithm:
For a formula \( f \), the program begins with the model \(< \{ w_0, f, 0 \}, \emptyset, \{ w_0 \}, \emptyset >\).

1. Searching for a non processed formula.
2. Applying transformation and simplification rules.
3. If there is no contradiction, and if there is a non processed formula, do (1) and (2).
4. If there is a non-normal world \( w \), and \( R(w_0, w) \), stop, otherwise do (5).
5. Build a non-normal world \( w \), with \( R(w_0, w) \), put \( f \) if \( \Box f \in w_0 \), put FALSE if there is a \( \Box \) formula in \( w \).

### The BF construction:
Let us point out a logical property:

**Property 1** Let \( f, g, h \) be propositional calculus formulas:

\[
BF(f, g, h) \text{ is a S7-model } \Leftrightarrow \begin{cases} 
(i) & f \text{ is satisfiable}, \\
(ii) & g \land h \text{ is satisfiable}, \\
(iii) & \neg f \land \neg g \land h \text{ is satisfiable}.
\end{cases}
\]

Principles of the demonstration:
A \( BF(f, g, h) \) can be considered as triplet \(< W, R, S >\), where \( W \) is a set of worlds, \( R \) a relation on the worlds, and \( S = \{(w, f)/w \in W\} \) a set of formulas index-linked by worlds of \( W \). To prove the \((\Rightarrow)\) direction, we just have to find a modal valuation function \( V \) such as \( V \Leftarrow 1 \) on \( S \), and such as \(< W, R, V >\) is a S7 model. The other direction is obvious.
From a logical point of view, the construction of a BF with known arguments \( f, g, h \) is systematic. But there is another way of considering the construction of a BF, still with a logical approach. We can re-interpret BF as an update (or revision) operator. This will be developed in the conclusion. So the difficulty does not lie in the formal construction, but in the way we use BF as a relevant musical representation.

### 3.2 Application of the model

- **U.S.T. as logical language elements**

  - **Possible worlds:** The possible worlds are points in musical time, stated in the listener’s mind. The U.S.T. give us the possibility to characterize a piece of music as a succession of musical moments. Therefore the M.I.M. have described two Haydn sonatas with U.S.T.. That will be our input material.

  - **Propositional variables:** The propositional variables are the basic vocabulary of the listener. Their meanings are various, sometimes not verbalisable. U.S.T. offer us a whole vocabulary of description and interpretation of musical perception phenomena, consistent with our choice of musical moments.

Each U.S.T. will be viewed as the conjunction of propositional variables, which are the components of the U.S.T. (table 2).

<table>
<thead>
<tr>
<th>duration</th>
<th>reiteration</th>
<th>phases</th>
<th>matter</th>
<th>direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>del.</td>
<td>not del.</td>
<td>with</td>
<td>without</td>
<td>one</td>
</tr>
<tr>
<td>l</td>
<td>-l</td>
<td>r</td>
<td>-r</td>
<td>p</td>
</tr>
</tbody>
</table>

**Table 2:** U.S.T. components as propositional variables
Example: (figure 2)

Figure 2: Haydn sonata, HOB XVI 50


One possible interpretation of this extract can be represented by the BF below: $BF(d, d, r \to \neg d)$ (figure 3)

<table>
<thead>
<tr>
<th>USTn₀</th>
<th>$w₀₀$</th>
<th>$(1) \ d$</th>
<th>$w₁₀$</th>
<th>$(1) \ d \land (r \to \neg d)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USTn₁</td>
<td>$w₀₁$</td>
<td>$(1) \ Dd$</td>
<td>$w₁₁$</td>
<td>$(1) \ D(d \land (r \to \neg d))$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USTn₂</td>
<td>$w₀₂$</td>
<td>$(1) \ \neg Dd \land \neg Dd$</td>
<td>$w₁₂$</td>
<td>$(1) \ \neg D(d \land (r \to \neg d)) \land \neg D(d \land (r \to \neg d))$</td>
</tr>
</tbody>
</table>

Fig. 3: $BF(d, d, r \to \neg d)$

Worlds $w_{ij}$ correspond to U.S.T. $i$. The formulated laws start at UST $n₁$, and concern iteration $r$ and direction $d$.

The first U.S.T. suggest to the listener a musical march, a direction ($w₀₀(1)$). This direction is in the second U.S.T. again, like an echo. Therefore, a musical law in $w₀₁$ is created. Just afterwards, this march is broken ($w₀₂$). The listener perceives the iteration of an element, and he understands that this iteration is the cause of the rupture.

Explanation of the BF construction
Let there be three arguments which are the laws formulated by the listener, and the $w₀₀$ U.S.T.; at first, the program checks property 1. Then, it develops a S7 model which is consistent with U.S.T. (U.S.T. viewed as the conjunction of propositional variables).
A U.S.T. \( n_2 \) must be found under three conditions:

(i) no contradiction between \( w_0 \), \( w_{12} \), \( w_{12} \) formulas and propositional variables of UST \( n_2 \).

(ii) same thing with the \( w_0 \) worlds and U.S.T. \( n_0 \).

(iii) same thing with the \( w_{11} \) worlds and U.S.T. \( n_0 \ldots n_1 \).

\[ BF(d,d,r \rightarrow -d); \]

Property 1 ok. Bivalence function:

\(<\text{moving forward, floating, ready to move}>\)

The ligne \( i \) is the UST \( n_{i} \).

\(<<0, "d">, 1, <"d", "d">, 2, <"("(d&: (r->-d))", 10, <"("(d&: (r->-d))"),

11, <"("(d&: (r->-d))", "("(d&: (r->-d))"),

12, <"("(d&: (r->-d))", "("(d&: (r->-d))"), 13, <"("(d&: (r->-d))"),

10, <"("(d&: (r->-d))"), 10, <"("(d&: (r->-d))"),

13, <"("(d&: (r->-d))"), 13, <"("(d&: (r->-d))"),

12, <"("(d&: (r->-d))"), 12, <"("(d&: (r->-d))"),

11, <"("(d&: (r->-d))"), 11, <"("(d&: (r->-d))"

}\}

### 3.3 Limits and future prospects

For our first simulation of the production by a listener of a musical piece “intelligent representation”, we have worked on a special cognitive mechanism, UNLL. When attempting the automation of the Kunst model, we have met some limits. Firstly, the S7 system is used essentially as a representation mode essentially; its deduction power as a logic system is not exploited. The deduction from UNLearning to Learning is fixed. But with the recent theory of non-monotonic logic, we can consider BF as an update/revision operator. A first idea is to consider the revision of a Knowledge Base (the representation of the piece) by a new formula (the new perception elements) according to a set of formulas (the musical competence of the listener). This way, the formalisation could produce an explicit representation of the dynamic process of memory during UNLL.

Secondly, there are several perception phenomena integrated into the model. The selection, represented by musical worlds, is one of them, as well as the constitution of a description vocabulary. These phenomena need to be formalized (with the help of psycho-cognitive elements) in order to build a machine capable to listen to music.

### References


L'INTELLIGENCE MUSICALE : ANALYSE ET PROCESSUS COGNITIFS.
Jean-Marc CHOUVEL (Université de Lille III)

Piaget résumait les deux facultés essentielles de l'intelligence comme fonction d'organisation et fonction d'adaptation. Il semble que la plupart des travaux analytiques visent à rendre compte d'un « système » plus ou moins clôt qui fixerait les actes d'attention au sein d'une organisation structurelle préalable. Comment, pourtant, rendre compte de la capacité, voire de l'appétit, pour la nouveauté, qui est le fait de bien plus de mélanomes que les cercles vicieux des sondages d'opinion et les a priori institutionnels voudraient le reconnaître? Il faut pour cela renouveler le modèle de l'expérience et du rapport au monde. En d'autre termes, le structuralisme, pensé pour un espace statique, bousculé par la temporalité des événements, cède la place au cognitivisme, dans la mesure où ce dernier est capable de rendre compte de l'évolution et de l'adaptation. Quelles sont alors les points de repères imprescriptibles sans lesquels aucune nouveauté ne peut advenir, c'est à dire prendre part à la constitution de l'acquis?

On tentera de répondre à cette question en revenant au formalisme de la théorie de l'information, et en montrant en particulier qu'un message quelconque peut être « décodé » en l'absence de la connaissance d'un code préalable, ou plutôt, en conséitant ce code au fur et à mesure des possibilités de la structuration du flux informationnel. Ce résultat théorique autorise une méthodologie analytique autonome vis à vis des présupposés culturels, et permet de saisir pourquoi le champ musical se situe à la fois en deçà et au delà du champ purement linguistique.

I. Structure et forme des processus temporels.

Etant donné la difficulté de définir la structure et la forme d'une pièce de musique, et la confusion souvent entretenu entre ces deux termes, tentons de préciser les démarches inhérentes à chacune de ces notions. En passant de l'objet à l'acte, nous pourrons peut-être mieux appréhender comment elles se transposent de l'espace à la temporalité. Qu'entendons-nous par structurer et par formaliser? Structurer, si l'on se fie à l'étymologie, aurait à voir avec construire; formaliser, si l'on en croit le formalisme, tiendrait plutôt d'une soumission à des conditions universelles a priori. En retrouvant la forme comme préalable et la structure comme déploiement, nous n'avons certes pas beaucoup avancé. L'hypothèse demeure que l'œuvre peut toujours être décrite comme un ensemble de constituants, un ensemble où se répartit une certaine substance, substance qui donne sa matérialité à l'œuvre. A ce niveau, peu nous importe d'être chez un architecte ou chez un musicien. Émpruntons quelques termes à la théorie des ensembles : le premier, c'est celui de relation. La première urgence de l'intelligence semble bien être de définir les liens internes entre les constituants, ou externes, des constituants avec d'autres références extérieures à l'œuvre. Ce lien peut être de nature différente de l'ensemble. Cette distinction, fondamentale, nous permet de mieux comprendre les mathématiciens. Après avoir parlé de relations, il parle d'ordre et d'équivalence. D'un côté, inclinations, successions, hiérarchies, de l'autre égalité, similitude, apparemment.

Le tableau suivant permet de mieux comprendre les différents aspects de ces distinctions qui pour très schématiques qu'elles soient n'en sont pas moins tout à fait opérantes. Il correspond toutefois à des approches où la temporalité n'a pas été intégrée. Structurer, dans l'établissement des liens internes, serait donc principalement donner un ordre, de la pierre à la clef de voûte, de la note à la symphonie. Que signifie alors « formaliser »? S'agirait-il de traiter des équivalences, entre deux piliers, deux colonnes, deux frises, entre deux notes, deux thèmes, deux textures? Peut-être cela, et au delà de la variation, chercher le modèle, au delà de l'objet, la catégorie, au delà de l'identique, le semblable. Reste le troisième lien, ce lien externe associatif, qui autorise les codes sémiotiques, et les analogies subjectives.

545
II. Elaboration de la structure.

Nous avions proposé\(^1\) de scinder le problème de la forme musicale en trois catégories: l'Espace, le Modèle et l'Objet. On conçoit aisément que chaque passage entre « ensemble d'objet » et « espace », exige le franchissement d'un niveau de la structure. L'idée même de ce franchissement permet de donner une interprétation tout à fait originale au problème du groupement et de la segmentation. Il semble dans la nature même de la temporalité d'imposer ce parcours du microcosme au macrocosme, du son à l'œuvre. On écoute la première seconde d'une œuvre avant d'avoir écouté la première minute, on entend la note avant la phrase mélodique... Cela est vrai au moins au début de l'œuvre, avant que la mémoire n'ait pu bouleverser notre rapport au temps. En effet, notre situation spontanée est très probablement intermédiaire: les structures de base sont innées, ou acquises\(^2\), alors que la compréhension des hautes structures nous demande un effort particulier. Les notes se regroupent sans effort en objet thématique alors que nous n'avons pas toujours pris conscience de la construction de la symphonie.

La relation d'inclusion est plus naturellement représentée dans l'espace que dans le temps, et il nous a fallu quelques années d'effort intellectuel avant de pouvoir en donner une représentation algorithmique convenable. Cette relation, qui ordonne l'idée de structure, nécessite un niveau zéro qui n'est pas forcément à confondre avec le fameux « niveau neutre » des sémiologues. Ce niveau zéro est en fait le niveau substantiel en deçà duquel la notion même de structure n'est plus opportune. C'est, pour reprendre l'idée grecque, un niveau « atomique », infragmentable. Tout ce que nous avons dit précédemment de la notion d'événement est bien entendu à rapprocher de ce niveau zéro, niveau qui reste malgré tout un choix, une hypothèse sur la résolution de notre appareil percutif. Aussi, tout objet répondant au critère de non pertinence d'une subdivision peut être pris, le cas échéant, comme élément du niveau zéro. Par exemple : l'événement des physiciens, l'échantillon, le contenu d'une fenêtre d'intégration dans le cadre de l'analyse de Fourier, l'événement MIDI, la note, l'accord, etc...

Le fait de décider qu'un ensemble d'événements forme une unité est une tâche passablement complexe, et les critères qui déterminent ce choix peuvent qui plus est entrer en conflit. La théorie classique de la phraséologie, énoncée par Rieman, donne un bon exemple de


\(^2\)Le passage du spectre à la note, par exemple, fait partie de ce que nous appelerrions l'inné, et qui est, d'une manière générale l'objet de la psychoacoustique. L'acquis tiendrait plus à des notions de type culturel, comme par exemple les formules cadentielles.
schéma fondateur. Un schéma de même nature a été repris et développé par Meyer sous la forme d'une méthode d'analyse rythmique. Lerdhal et Jackendorf ont proposé quant à eux des règles préférentielles de groupement plus générales, tout en restant dans le cadre de la musique tonale. Le fait de regrouper des éléments n'est pas a priori limité. Simplement, il est bien certain que tous les groupements ne sont pas « cognitivement corrects ». Sans rentrer dans la description de cette qualité probablement très variable selon les œuvres et les individus, nous pouvons toutefois tenter une description algorithmique minimale du mécanisme de structuration, description qui fait de toute façon apparaître la nécessité d'un certain nombre d'opérations et de tests.

```
niveau := 0; ...

pour tous les événements élémentaires (i de 1 à n):

objet[niveau, i] := événement-élémentaire[i];

(acquisition du ou des événement(s) en cours)

structuration[niveau, objet[niveau, i]]:

{objet-hypothétique[niveau+1, i] :=
objet-hypothétique[niveau+1, i] "U" object[niveau, i];

(intégration)

si objet-hypothétique[niveau+1, i] est achevé :

(les critères formels d'achèvement — de complétude — sont atteints)

{object[niveau+1, i] := objet-hypothétique[niveau+1, i];

(mémorisation de la définition structurelle de l'objet)

niveau:= niveau+ 1;

structuration[niveau, objet[niveau, i]];

niveau:= niveau- 1;)

(mécanisme récursif - changement de niveau)

(définition de la fonction de structuration)
```

Cette présentation algorithmique est évidemment assez élémentaire, mais elle donne dès à présent un cadre à nos réflexions. Ce mécanisme réalise en effet en temps réel une opération de structuration des événements dont il est saisi. Il le fait à la condition d'une part de pouvoir réaliser une opération que nous appelerez « d'intégration » (représentée par "U" dans l'algorithme), et d'autre part de disposer d'un critère de complétude qui lui permette de fermer les groupements pour accéder à des objets de niveau supérieur. Notons également qu'il a besoin d'une mémoire de travail, ou d'une mémoire temporaire pour pouvoir exercer l'ensemble de ces opérations et de ces jugements. Le mécanisme récursif est ici une pure commodité d'écriture, et éventuellement d'implémentation : les critères évoqués précédemment n'ont a priori pas de raison d'être identiques à chaque niveau. En fait, même si on pourra donner un exemple théorique d'une telle réalisation, la réalité cognitive humaine semble nous orienter vers une spécialisation des fonctions, et cela d'autant plus que le niveau est élémentaire. Ainsi, la physiologie de notre oreille nous donne directement accès à la fréquence et au timbre de la note, sans nous encombrer de la multitude de périodes qui y contribuent. De même, nous appréhendons globalement les mots du langage sans nous arrêter à chaque syllabe. Peut-être en est-il encore de même pour le rapport des mots à la phrase. C'est évidemment de moins en moins vrai quand on atteint les grandes articulations du discours. Notons au passage que l'art lutte souvent frontalement contre ces évidences. C'est le cas par exemple pour la poésie, ou la valeur mélodico-timbrique des syllabes peut être parfois prépondérante sur la valeur sémantique des mots.

### III. Elaboration de la forme.

Nous avions depuis longtemps explicité les mécanismes de réalisation temporelle de la notion de forme. Sans doute, de par la prééminence historique du type d'analyse correspondant à la notion d'équivalence (analyse motivique, analyse paradigmatique...), cette explicitation venait comme un prolongement d'une logique déjà mise en place. Il convient toutefois de

---


redonner ici l'algorithme initial, réécrit avec les notations précédentes, en faisant apparaître ainsi les interactions que les deux algorithmes peuvent entretenir l'un avec l'autre.

\[ \text{niveau} := 
\]

\text{pour tous les objets[niveau ,i]} :

\text{si objet[niveau ,i] ne sont pas semblables aux objets présents dans la mémoire matériau :}

\{\text{écrire objet[niveau ,i] en mémoire-matériaux} \}; \text{(constitution de la mémoire matériau)}

\text{garder une trace de la position de objet[niveau ,i] dans la mémoire-forme ;}

\text{(constitution de la mémoire forme)}

Une des conditions essentielles de la possibilité d'un test de similitude réside dans le fait que pour pouvoir être comparés, deux objets doivent appartenir au même niveau structurel... Le test de similitude peut être forcé en un test binaire d'identité, mais des degrés de variation peuvent aussi être pris en compte dans le cadre de ce que nous avions appelé des « matrices de présence », qui sont une présentation plus complexe de la mémoire forme permettant d'accéder à une notion plus floue (et plus subtile) du « modèle ».

Ce qui est fondamental dans les propositions précédentes, c'est de bien comprendre à quel point les deux notions de forme et de structure, conceptuellement parfaitement séparées, ont besoin l'une de l'autre pour pouvoir être opérantes. La structure nécessite la forme pour définir les modèles de ses objets et pour ainsi assurer la justesse de ses groupements ; la forme a besoin de la structure pour rendre possible sur un niveau cohérent ses associations et pour ainsi donner un sens à la notion de similitude. Qui plus est, forme et structure ne désignent certainement pas des processus cognitifs séparés, et il nous faut maintenant imaginer leurs interactions.

IV. Comment appréhender la structure et la forme d'un flux d'événements : proposition pour un modèle cognitif global.

La mémoire est un élément déterminant de toute cette étude, ceci non seulement du fait de la nécessité de son intervention, mais aussi et surtout de par cette contrainte, ou pour être plus exact cette hypothèse de contrainte, qui oriente les mécanismes cognitifs vers ce que nous appellerons une « efficacité mémorielle ». En quoi consiste cette efficacité? Et bien principalement dans la rétention du maximum d'information avec le minimum de moyens. Autrement dit, le meilleur mécanisme cognitif serait celui qui retient le plus en mémorisant le moins. Si l'on pousse l'image, c'est celui qui permet de restituer intégralement ce qui a été perçu en utilisant le minimum d'espace-mémoire. Nous savons d'ores et déjà qu'en réalité, notre propre intelligence opère de nécessaires compromis. Mais le propos est ici de comprendre ce que serait ce comportement idéal, en sachant qu'il peut tout à fait servir de modèle fondamental auquel des modèles plus réalistes pourront être comparés. Il va de soi qu'un tel modèle est là pour dégager des concepts de référence, pour la plupart communs au versant analytique et au versant psychologique de notre entreprise. Il doit d'autre part être opérationnel pour l'analyse, en dehors de tout souci de réalisme psychologique.

La mémoire intervient, comme nous l'avons dit, dans la structuration, en tant que mémoire temporaire tant que l'objet hypothétique n'a pas été intégré en tant qu'objet. La mémoire « statique » nécessitée par la forme doit permettre au moins la mémorisation des événements élémentaires. La forme et la structure coïncident dans le fait qu'un objet d'un niveau donné est défini par ses relations aux objets qui composent, le niveau directement inférieur. Cela signifie que le processus que nous sommes en train d'envisager ne devra mémoriser, en plus des événements élémentaires, que ces relations de niveau à niveau, et cela uniquement dans la partie du processus qui correspond à une découverte et à tous les niveaux jusqu'au niveau final de l'œuvre entière. Pour illustrer ce mécanisme, la figure suivante reprend une analyse du dernier mouvement allegretto de la sonate de Mozart K. 545. Pour simplifier cet exemple, on s'est contenté d'événements élémentaires basés sur une segmentation systématique toutes les deux mesures. La structure a été représentée avec les conventions de la théorie des ensembles. Les objets représentés en lignes pointillées ne font pas partie du front de découverte.
du niveau correspondant, et donc ne nécessitent pas d'être mémorisés à nouveau. Les parties noires représentent donc le matériau et les liens absolument nécessaires à une représentation intégrale du mouvement.

**Fig. 2 : Représentation formelle et structurelle de l'Allegro de la sonate K. 545. de W. A. Mozart**

Qui plus est, outre le fait de ne pas engager la mémorisation dans le sens de l'intégration, les zones grises peuvent au contraire donner lieu à une anticipation, c'est à dire justifier d'un mécanisme cognitif tout à fait différent qui permet de prévoir la suite dès que le début a été reconnu.

**Fig. 3. : Représentation de la structure donnant un sens temporel aux liens de regroupement.**

En fait, il faut encore distinguer deux cas : celui où cette prévision s'avère fondée et celui où la suite se révèle différente. C'est ce que permet de visualiser la représentation précédente qui reprend, en donnant un sens à la situation temporelle des nœuds de groupement structuré, la représentation traditionnelle « en arbre ». Dans le processus de réception de l'information l'esprit est engagé dans des phases assez différentes dont le mode de succession n'est sans doute pas indifférent dans la production d'un sens esthétique ou d'une émotion. On pense tout naturellement, au vu de la discussion précédente, au modèle d'implication-réalisation de L. B. Meyer. Il faut cependant introduire un certain nombre de nuances. D'abord, le système considéré n'est pas le même : dans l'exemple abordé ici, il s'agit d'une pièce considérée comme un tout autonome pour une analyse de type interne, chez Meyer, le système intègre l'ensemble de la sphère culturelle. D'autre part, l'analyse repose sur un niveau informationnel autonome, alors que Meyer fait clairement allusion à des normes esthétiques et culturelles. Néanmoins, la réflexion de Meyer nous rappelle que le monde psychique ne

---

commence pas au début d'une œuvre musicale, et que dans une représentation plus exhaustive, ces connaissances préalables doivent être prises en compte.

Nous pouvons désormais proposer une représentation générale qui inclue tous les éléments précédents, sous la forme d'un algorithme fonctionnel capable de rendre compte des trois types de liens que nous avons énumérés.

![Diagramme de représentation cognitive globale](image)

**Fig. 4. : Représentation cognitive globale.**

**Conclusion**

La simplicité apparente de ce schéma ne doit pas nous leurrer sur la complexité des réalités qu'il recouvre et qu'on ne peut ici que très brièvement évoquer. On peut bien entendu le prendre au pied de la lettre et considérer, comme nous l'avons fait dans notre exemple, qu'aucune mémorisation préalable n'est à prendre en compte, que la similitude peut se réduire à l'identité pure et simple et que la complétude est décidée à un niveau donné dès qu'un retour à un objet déjà connu est constaté au niveau inférieur. Ces règles extrêmement sommaires peuvent être suffisantes dans certains cas, et même servir de point de départ analytique. Mais d'une manière générale, la « similitude avec des objets antérieurs » et l'« achèvement » doivent être compris comme le résultat de processus et de décisions cognitifs complexes, mettant en jeu, en particulier du fait de l'immensité des connaissances préalables potentielles chez certains individus, des « bases de données » considérables, accessibles quasi instantanément. Le graphisme de la représentation précédente emprunté aux conventions de l'intelligence artificielle ne préjuge en rien de la nécessité de réseaux neuraux complexes pour modéliser vraisemblablement certaines fonctions. De plus, tout cela n'est que la représentation d'une réalité plongée dans le temps pour laquelle la performance du traitement du flux des événements est un facteur limitatif considérable...

**Bibliographie :**

A Cognitive Approach to Musical Analysis:  
Metaphorical Projection in Music  

Hallgjerd Aksnes, University of Oslo, Norway

In the following I will present a listener-oriented analytical approach that seeks to reflect the heterogeneity and embodiment of musical experience. The objects of analysis will thus be sought not in the score or in the "music itself", but rather in the listeners' cognitive processing of sounding music and score - a processing which involves both auditory, visual, emotional, kinaesthetic, linguistic, and other domains of experience. The notion of metaphorical projection will be employed as a tool for analyzing these processes; metaphor being defined in accordance with cognitive semantics as a bodily motivated "cross-domain mapping in the conceptual system" (Lakoff, 1993: 203). I wish to argue that many of the cross-modal associations we make during music listening are the result of metaphorical projections between different cognitive domains; and that metaphorical projections are so pervasive in musical experience that we tend to regard the most conventional metaphors as intrinsic to the "music itself" (e.g. the ubiquitous metaphor of musical space and the convention of describing timbre in terms of color).  

Although there is mounting evidence that metaphorical projection is a cognitive reality in many different domains of experience (Lakoff & Johnson, 1980; Johnson, 1987; Lakoff, 1987; Gibbs, 1994; Turner, 1996), the use of metaphors in musical discourse - especially metaphors with strong emotional connotations - has often been rejected as hopelessly subjective and unscientific by music analysts. However, as the music theorist Marion Guck has demonstrated, even the most scientifically oriented analyses abound in metaphorical language (Guck, 1994). Like Guck, I wish to argue that many of the metaphors which arise through our encounters with music can in fact be intersubjective, due to shared biological and cultural dispositions among the listeners. To demonstrate my point, I will start out by playing the Norwegian composer Geirr Tveitt's "O be ye most heartily welcome" from the orchestral suite A Hundred Folk Tunes from Hardanger (1954-63):

CD: Simax PSC 3108, Geirr Tveitt: A Hundred Folk Tunes from Hardanger, Suite no. 1

I have chosen to focus upon a section towards the end of the piece, as this can exemplify two contrasting metaphors which, I believe, are characteristic of the piece as a whole and contribute greatly to its import. The first and dominant metaphor is that of floating in a seemingly weightless state (cf. overhead, section E), whereas the second metaphor is that of gravitating or being pulled towards something (cf. section D):

OVERHEAD: Geirr Tveitt: "O be ye most heartily welcome", mm. 35-41.

How, then, can such associations arise? Let us first look at the "gravitational" part of this example. The measure preceding m. 35 ends on a dominant chord - thus, m. 35 starts as a deceptive cadence, but soon moves by way of an appoggiatura down to the tonic e aeolian. The movement does not stop here, however: The bass line continues downwards, reaching a c in m. 36.

1 Research funded by NFR, the Norwegian Research Council.
2 Interestingly, recent neurological findings support the view that music perception is cross-modal. Robert Zatorre and his associates have observed that passive listening to unfamiliar tonal melodies entails a cerebral blood flow (CBF) increase not only in the areas of the brain that are known to be specialized for auditory processing but also in the visual cortex; more precisely in the right occipital cortex (fusiform gyrus). Zatorre et al. also report an insignificant CBF increase within area 19 (extrastriate visual cortex) and refer to similar findings by Griffiths & Brown, who have raised the possibility that their findings reflect a cross-modal spatial perceptual system engaged by the apparent movement of frequency-modulated tones (Zatorre et al.: 1994). However, it remains to be seen which parts of the brain are active during an aesthetically and emotionally engaged listening experience.
3 Kinaesthetic metaphors such as these are pervasive in musical experience, as we tend to hear music in terms of agency (see Guck, 1981; 1991). However, as the music theorist Fred Everett Maus has observed, we find a pervasive indeterminacy in the identification of musical agents (Maus: 68).
36, which defines a sixth chord on the fourth scale degree; and thereafter a b, which is sustained as a dominant pedal point through m. 38. This last measure is an almost exact repetition of m. 37 – thus, it delays our arrival at the longed-for tonic, thereby greatly increasing the expressivity of this passage.

But how is it that the tonic can be heard as a point of arrival, and how is it that we can actually long for the tonic? In an attempt to understand this, I will turn to the philosopher Mark Johnson, who has postulated a number of basic conceptual structures which he calls “image schemata” and describes as “recurring structures of, or in, our perceptual interactions, bodily experiences, and cognitive operations” (Johnson: 79). According to Johnson, these basic structures can be metaphorically projected onto many different domains of physical and psychological experience; thereby enabling associations to be made between phenomena which pertain to different cognitive domains. Among the image schemata which Johnson discusses, we find the PATH schema which, as he observes, is often linked with the metaphor STATES ARE LOCATIONS (Ibid.: 114). We use these structures in many different ways, and especially as a way of creating temporal order – like we do when we hear the arrival of the tonic as a location towards which the music strives, thereby drawing a path. But even more interesting than this ubiquitous and very obvious metaphorical mapping, is the mapping which enables us to “long for” the tonic in mm. 37-38. Here we are acting according to a cultural convention which is limited to the paradigm of modern tonal harmony. However, although this musical effect is culturally determined, it seems to be motivated by our bodily being in the world. In order to understand how this is possible, I will turn to what Johnson calls the ATTRACTION schema:

A magnet draws a piece of steel towards itself, a vacuum cleaner pulls dirt into itself, and the earth pulls us back down when we jump. There is a common schematic structure of attraction shared by these experiences. This same structure is present, too, when we feel ourselves physically attracted to some other person. The force is not gravitational, in the standard sense, but it is a kind of gravitation toward an object. As such, it shares the same underlying ATTRACTION schema. (Ibid.: 48)

The strong “yearning” effect of the dominant pedal point in our example seems clearly to be based upon this schema, the tonic acting as a musical center of gravity. Interestingly, the “father” of modern tonal harmony, Jean-Philippe Rameau, was a contemporary of Isaac Newton; and Rameau draws explicitly upon Newton’s theory of gravity in his own theory of harmony (see Christensen: 189). To this day, we experience harmony in gravitational terms: The main driving force of functional harmony is the tension between the dominant and the tonic, as epitomized by the dominant chord’s leading note being pulled towards the tonic. Furthermore, we both hear and feel this harmonic “suction” as a major source of musical tensions and releases – and accordingly, as an important expressive means in tonal music.

Returning now to our musical example, let us take a closer look at section E. As we can see from the score, the tonic is reached in m. 39, but this cadential movement does not create a very strong sense of closure and rest. Instead, the music seems to start floating, as it did in the beginning of the piece. How can this sensation arise? Let us again take the structural features as our point of departure. First of all, the bass line does not make a standard cadential move in m. 39 but continues to lie on the b, thereby defining a tonic six-four. Tonic six-fours are rarely found on strong beats and in cadential passages, since they are conventionally heard as relatively unstable. When presented as passing notes tonic six-fours can contribute to the directional drive of a musical passage – but when they are presented as in this example, without being followed by a more stable musical structure, the effect will be just the opposite: the music will tend to lose its directionality. In addition to this voice-leading effect in mm. 38-39, the subsequent rotation between six-four chords on the first and second scale degree also contributes to the floating character of this passage; and the static harmonic effect is enhanced even further by the two chords’ common notes, e and b (the first and fifth scale degree of e aeolian). To summarize: Tveitt has substituted the linear, functional harmonic concept of section D with a more static, coloristic harmonic concept in section E, thereby breaking the convention of harmonic development (cf. French impressionism).
The lack of development in section E is not limited to the harmonic structure, however; in rhythmic and melodic respects, too, this passage is strikingly static. To understand why, let us take a closer look at its composition. The passage has a very simple structure, consisting of three sonorous layers: sustained chords in the strings, repeated quarter notes in the harp, and the slowly pacing folk tune presented by a clarinet. The changes in each of the layers are slow and very subtle; even in the folk tune, which consists solely of two repeated, closely related rhythmic motives. The melody revolves mostly around the tonic, and with but one exception, it moves only stepwise or by thirds employed in eighth note rotations between the first and sixth scale degree (which occurs twice in the folk tune). This structural stasis contributes greatly to the floating sensation of the passage—a sensation which is enhanced even further by its slow tempo and subdued, static dynamics.

If we now analyze this analysis, we will discover that we are again experiencing the music in terms of the PATH and ATTRACTION schemata—this time, however, in the form of a negation. Due to the harmonic, rhythmic, melodic, and dynamic peculiarities of this passage, the music seems no longer to be progressing; and it seems to have liberated itself from the gravitational forces to which we earthly beings are subject. This has dramatic temporal consequences, as it seems that time itself starts to float. Another way of expressing this would be to say that the music has an ethereal character—a metaphor that is based upon our experience of things floating in air, as well as upon our cultural images of an incorporeal and atemporal afterlife. These metaphorical associations are enhanced by the ethereal connotations of the instrumentation, where we find a dominance of muted strings, harp, and soft-sounding wind instruments; all of which figure prominently in the "heavenly" musics of our culture.

I would have liked to study the folk tune more closely in terms of another important image schema, the BALANCE schema—but due to the time limit, I have chosen instead to turn to an aspect of musical meaning which seems to evade the more kinaesthetically oriented image schemata. Johnson's schemata can help us to understand how we draw upon our own experiences as moving bodies when we listen to music; but they are less apt for understanding many of the subtle dynamic effects which are so important for musical expressivity. The developmental psychologist Daniel Stern, however, has introduced a notion which can account for just such effects: the notion of vitality affects. Stern describes vitality affects as amodal activation contours which manifest themselves in many different experiential domains, and which act as a complement to the Darwinian categorical affects:

...many qualities of feeling that occur do not fit into our existing lexicon or taxonomy of affects. These elusive qualities are better captured by dynamic, kinetic terms, such as "surging", "fading away", "fleeting", "explosive", "crescendo", "decrescendo", "bursting", "drewn out", and so on. (Stern: 54)

Stern draws upon the philosopher Susanne Langer's discussion of the different "forms of feeling" that are correlated with vital life processes like breathing or the coming and going of emotions and thoughts, and he emphasizes that vitality affects can exist independently of the traditional categorical affects—"For example, a 'rush' of anger or of joy, a perceived flooding of light, an accelerating sequence of thoughts, an unmeasurable wave of feeling evoked by music, and a shot of narcotics can all feel like 'rushes'." (Ibid.: 55) He emphasizes further that the vitality affects are inherently expressive; a point he makes by pointing to the expressivity of abstract dance and music. In his discussion of a metaphor in Defoe's novel Moll Flanders, Stern provides a partial answer to why this is so: "...a variety of diverse sensory experiences with similar activation contours can be yoked—that is, they can be experienced as correspondent and thereby as creating organization." (Ibid.: 58) How is this possible? Stern explains:

Because activation contours (such as "rushes" of thought, feeling, or action) can apply to any kind of behavior or sentence, an activation contour can be abstracted from one kind of behavior and can exist in

---

4 I am grateful to Mark Johnson for having introduced me to the work of Daniel Stern.
5 Interestingly, the neuroscientist Antonio Damasio (private communication) has compared Stern's vitality affects with his own notion of background feelings; feelings which originate "in 'background' body states rather than in emotional states" (Damasio: 150; emphasis added).
some amodal form so that it can apply to another kind of overt behavior or mental process. These abstract representations may then permit intermodal correspondences to be made between similar activation contours expressed in diverse behavioral manifestations. (Ibid.: 57-58)

If we now return to Tveitt’s “O ye most heartily welcome”, we will recognize a number of vitality affects which contribute to the expressivity of this piece – and I am convinced that these elements’ expressivity is due at least in part to mappings between dynamic patterns in the music and similar patterns in our visceral experience. Let us start with the vitality affect “drawn out”, which is represented in two respects in our example. Firstly, it is represented by means of the repetitions in mm. 37-38, which, as we noted, entail a tantalizing delay of the tonic; tantalizing precisely because we react to this delay in the same way that we react to the delay of an expected gratification.6 The second way in which our example is “drawn out” has to do with the slow tempo: It is as if time is drawn out in this passage, as if we are experiencing the musical movements in slow motion – the same way that we can experience dreams, distant memories, and things sounding from far away. Thus, adjectives like “dream-like” and “distant-sounding”, which are often used to describe music with a character similar to this piece, can be regarded as metaphorical entailments of the vitality affect “drawn out”.

Among other vitality affects which are especially prominent in this piece, I wish to mention “surging" and “crescendo”, which play a vital role in the passage immediately preceding our example (mm. 31-34; CD 2.16-2.33). Here we find a dynamic crescendo effect which is combined with an octave leap in the deep strings, marking the repetition of a melodic motive that revolves around the fifth scale degree. The second time round the strings reach the tonic – only to fall back upon the leading note on the last beat of m. 34. As we can hear, this passage is characterized not only by increased loudness but by a general increase in intensity, due to the combined forces of dynamics, pitch, and harmonic expectations. The effect of this musical surging motion is enhanced even further by the following cesura; a vitality affect which, incidentally, is not marked in the score but is a gift from the conductor, Per Dreier. All of these musical effects are analogous in some way to our experience or expressions of emotional arousal and intense anticipation (as to the cesura, cf. the idiom “My heart stood still.”)

Now turning to our ability to hear music in terms of categorical affects, it is interesting to note that most attempts to explain emotional expression in music draw upon vitality affects as a basis for the perception of categorical affects. Susanne Langer (1951; 1953) who, as we have seen, is one of Stern’s inspirators, believes that our ability to perceive emotion in music reflects an isomorphism between the sounding music and our feelings. This view still has considerable influence today, as we can see from Dowling & Harwood’s Music Cognition, where emotional effects that depend upon patterns within the music itself are discussed in terms of iconic representations. Like Langer, Dowling & Harwood regard the formal similarities between music and emotion as a key to the understanding of the emotional effects of music, noting: “the ebb and flow of tensions and relaxations in the music mirror the form of emotional tensions and relaxations.” (Dowling & Harwood: 205) Dowling & Harwood also put great emphasis on the expressive quality of musical motion, citing Aristotle with approval on this matter: “...we feel the motion which follows sound. ... These motions stimulate action, and this action is the sign of feeling.” (Ibid.: 206) Interestingly, as Dowling & Harwood note, Aristotle here combines indexical and iconic representations, pointing not only to the iconic relationship between tonal motion and action but also to the prior indexical associations of actions and feelings (Loc.cit.)

It is precisely such indexically motivated metaphorical mappings that will be the objects of study in my next example – another piece from Tveitt’s A Hundred Folk Tunes from Hardanger (I will not give you the title until later, as I wish to keep the semantic connotations out of the picture for the time being):

CASSETTE: “Friarføter” from A Hundred Folk Tunes from Hardanger, Suite no. 4

6 Interestingly, the music theorist Leonard B. Meyer believes that the objectification of musical meaning is dependent upon precisely such delays, or inhibitions, of tendencies or habit reactions (Meyer: 39).
The piece as performed here, by Bjarte Engeset and the Bergen Philharmonic Orchestra, has many of the prototypical characteristics of music which is perceived to express happiness — to quote a recent study by Gabrielsson & Juslin: “fast tempo, moderate variations in timing, moderate to loud sound level, tendency to (relatively) sharpen contrasts between ‘long’ and ‘short’ notes (as in dotted patterns), [...] rapid tone onsets, bright timbre [...]” (Gabrielsson & Juslin: 86). In addition to these performance-related features, we can point to several of the music-structural features of “happy” music reported by Rigg (1964); such as the piece’s relatively high pitch, its high degree of consonance, and its G-lydian tonality (Lydiian is commonly heard to be even brighter than major, due to the raised fourth which acts as an upward-striving leading note.) To understand how it is that we can experience this piece and the above-mentioned musical features as “happy”, I will draw upon two experiential domains which, I believe, play an active role in our experience of the piece. One experiential domain has to do with our bodily expressions of happiness and the other with our bodily experience of happiness:

First of all, we commonly hear music in terms of manners of walking (see Swanwick: 29); and the Tveitt piece is no exception, due not least to the rhythmic ostinato figure which accompanies the folk tune throughout the piece. The quick, skipping gait of this ostinato can be heard as symptomatic of happiness, and the several unexpected, staccato, and high-pitched sforzato effects can be metaphorically conceived of as “jumping for joy” (Note that this does not imply that the sforzato effect cannot also have specifically sonorous meanings.) The analogy of walking is but one possibility, however. Charlotte Wolff’s observation of our postural and gestural expressions of elation can also throw light upon the workings of the Tveitt piece, as every one of Wolff’s observations can be recognized in this piece: “Elation is shown by a wealth of unnecessary movement, fast motor speeds, exhibitionist behaviour, spontaneous, emphatic and rhythmical gesture and self-assertiveness.” (Swanwick: 28) Returning now to Gabrielsson & Juslin’s study of the emotional expression of happiness in music performance, we find several striking correspondences with Wolff’s study; correspondences which indicate that the perception of emotional expression in music is motivated to a large extent by our bodily being in the world. The Tveitt piece and Gabrielsson & Juslin’s findings correspond not only with our motor expressions of happiness, however; they also correspond with our vocal expressions of happiness (cf. especially the piece’s moderate to loud sound level and bright timbre, as well as its relatively high pitch) — which represent yet another layer of the tightly woven fabric of musical meaning.

But what about the second domain, the domain of our bodily experience of happiness? Interestingly, many of the same musical features which could be heard in terms of our bodily expressions of joy can also be heard in terms of our experience of joy. The quick tempo is in accordance with the heartbeat of an elated person, the relatively high pitch can be heard as indicative of “high spirits” (another metaphor which is based upon our postural expression of happiness), and the sforzato effects can be metaphorically conceived of as a heart (my heart?) “making a leap”. Furthermore, the accelerando in the opening of the piece can be heard as indicative of excitement and anticipation (cf. the physiological effects of emotional arousal); and the sudden dynamic swell as a “rush” of joy. It should be noted that we rarely tend to objectify such associations during the fleeting music listening experience — many of the associations emerge only as we stop to reflect upon this experience. However, it is plausible that cross-modal associations of this sort play an important role on the subconscious levels of the music listening experience, as they contribute to the network of associations which constitutes the emergent meaning at each moment of this experience.

As we have seen, one and the same musical feature can be heard as indicative of several different phenomena; a peculiarity that is due to the relatively indeterminate nature of musical meaning (see Maus, 1988 and Walton, 1994). However, the associations that I have discussed are far from being arbitrary, as the phenomena at issue can be characterized by similar rhythms, movements, forces, or patterns of agency — the very same aspects of music which have been discussed in terms of image schemata, vitality affects, and forms of feeling. Furthermore, it turns out that this particular folk tune favors precisely the associations upon which I have drawn: The folk tune is titled “Friarfôster” which, literally translated, means “Wooer’s Feet”, alluding to the Norwegian idiom “å være på frierfôster” — “to be on wooer’s feet”. And if we
wish to semanticize the piece even further, we can examine the folk tune’s original text: “Bottolf was unable to travel all the way to Røldal to ask for Anna’s hand in marriage, so he sent her a message. When she received his proposal, she sent him a spoon. Her suitor was delighted at such willing acceptance.” (Storaas: 27) Geirr Tveitt often sought musical inspiration inextramusical phenomena, as we can see from the composer’s own statements regarding A Hundred Folk Tunes from Hardanger, as well as from the many programmatic elements in his music (see Olsen, 1970). It is therefore plausible that this particular folk tune has elicited a vast amount of associations pertaining to love, happiness, excitement, anticipation, hope, and joyfully skipping feet in Tveitt’s musical imagination. And as I hope to have demonstrated above, we can recognize many of these associations even without access to the title and textual basis, due to our capacity for metaphorical projection.

The aim of this presentation has been to use the notion of metaphorical projection as a music-analytical tool which can help us to continue from where the traditional analytical methods stop; the underlying conviction being that the allegedly autonomous musical structures of traditional analysis represent but one aspect of our rich, complex, and highly heterogeneous musical experiences. What I hope to have demonstrated above, is that even metaphors which have wandered far from the domains that are commonly included in so-called structural analysis — and which are often dismissed as too subjective or idiosyncratic to be of any analytical use — can be highly communicable, due to our shared biological and cultural worlds.

References


Chunking in music theory by imagined sound-producing actions

Rolf Inge Godøy
Section for musicology
University of Oslo
P.O.Box 1017 Blindern
N-0315 Norway
tel. (+47)22854064, fax. (+47)22854763
E-mail: r.i.godoy@imt.uio.no

1. Introduction

The aim of this paper is to show how various imagined sound-producing actions (hitting, bowing, blowing, kicking, etc., singly or in combinations) can enhance our understanding of chunking and coherence in musical objects, giving us clear and vivid images of segmentations, groupings (motives, phrases, textures, contours, etc.) and hierarchies in the analysis of music. Chunking is here not only a matter of parsing auditory streams, but equally a matter of transforming the continuous auditory stream to more or less discontinuous images of musical objects, what could be called a flux to solid transformation in our cognition. This is based on the assumption that transformations from continuous acoustic input to some kinds of discontinuous meaning units are always going on in listening, and that the production of chunks or overview images of musical sound is in fact the very condition for discrimination and knowledge in music in the first place. However, this "compression" of continuous acoustic flux into more stable and solid images in our memory is still enigmatic, and has been the subject of phenomenological and gestalt research in the past as well as of various contemporary cognitive sciences. Available material indicates that this flux to solid transformation in musical cognition depends upon both an analysis of the acoustic signal and on schematic orderings based on previous learning. Experience of various sound-producing actions, as well as other bodily based schematic images, could be influential in this chunking process, and I shall try to give a brief review of some material in support of this as well as some ideas for further research.

Another, but until recently mostly neglected issue here, is musical imagery as an integral element of music theory. Mainstream music theory in our culture has focused on symbol-based abstractions, ignoring the issues of our images of musical sound. With the notable exceptions of some previous phenomenological and gestalt research, musical imagery was not really put on the agenda of music theory before Pierre Schaeffers project of "questioning the listening consciousness" (Schaeffer 1966, p. 147). Schaeffers strategy with the musical object can be summarized as that of contemplating images of fragments of musical sound, progressively discovering and mapping out more and more features and ordering these features into a multi-
dimensional morphological model of the musical object. Fortunately, there has recently been increasing interest in auditory imagery (Reisberg 1992), however this field is much less developed than that of visual mental imagery (Kosslyn 1994). Of particular interest here, research seems to indicate that there is a need for "stimulus support" in auditory imagery (Smith, Reisberg and Wilson 1992), meaning that images of sound-producing actions are involved in the recall and contemplation of auditory images. For the moment, we could say there are two main reasons for focusing on imagined sound-producing actions in music theory:

A) Sound-producing actions are necessary for musical imagery as stimulus support, and hence for representations of musical objects in our thinking.

B) Sound-producing actions are a contributing schematic factor in transforming continuous acoustic flux to more stable objects or entities of meaning in our consciousness.

We have in recent years seen the emergence of more "ecologically" founded projects in music theory, based on biological and environmental constraints for auditory cognition as well as on the complex, sub-symbolic substrates of musical sound, rather than on the symbolic abstractions of conventional western music theory. This includes the large field of research in digital synthesis and signal processing (Moore 1990, Risset 1991), with work which attempts to extract patterns from this complex substrate of musical sound in a bottom-up manner (Ellis 1995), as well as work which takes into account both self-organization in the signal and in the modeling of the ensuing cognitive process (Leman 1995). Also, there is of course the substantial body of audition research (Bregman 1990, McAdams and Bigand 1993), which constitutes a broad basis for talking about schematic integration of acoustic signals in audition.

From the available material, I would like to focus on schematic integration by sound-producing actions, as I believe that a purely "signal based" approach is not sufficient to understand how images of musical objects are formed. Auditory scene analysis, or listening in general, seems to call upon previous experience of source identification, and this source identification can even override incoherent or degraded acoustic signals (Bregman 1990). I believe we have good reason to suspect that there is a strong top-down schematism by imagined sound-producing actions in auditory imagery as well, and for this reason the kinematic images of sound production (as well as other body schematic images) could be seen as integral to the musical objects in our imagination. Drawing the shapes of action trajectories, or the silent choreography of sound-producing actions, will for this reason actually become a part of musical objects, making it necessary to consider some issues of cross-modality.
2. Cross-modality

Although the subject of cross-modality raises difficult questions about the criteria for defining distinct modalities and about the interaction of various modalities in cognition, I think we here can adopt the simplified approach of considering the acoustic signal, action and visual images as three distinct modalities, and consider their interaction as a matter of concurrence and mutual activation. In this simplified sense, any association of sound with sound-producing action or any association of visual shape with the trajectory of a sound (such as its envelope) or its associated trajectory of sound-producing actions (its kinematic image), could be considered cross-modal. This view of cross-modality (Stein and Meredith 1993) could be seen as related to analogical thinking (Holyoak and Thagard 1995), or in general to the use of metaphor (Johnson 1987, Lakoff 1987). It is well known that visual metaphors are abundant in musical discourse, where terms such as "high", "low", "long", "short", "thick", "thin", etc. are quite indispensable even for the most abstract projects of music theory. As I have tried to argue elsewhere (Godøy 1996), and as has been admirable demonstrated by Schaeffer (Schaeffer 1996, Chion 1983), the extensive use of visual metaphors can be fruitful in music theory explorations, creating awareness and discriminations of qualities in musical sound which previously has not been thematized by western music theory, in particular qualities of timbre and texture.

However, the actual workings of cross-modality in musical imagery is more enigmatic and not well explored. The intuitions of the early gestaltists pointed in the direction of musical cognition and imagery as holistic processes in the sense of calling upon the cooperation of large areas or fields of experience in memory. As far as I can see, some contemporary research seems to agree with the introspections of the early gestaltists, indicating close links between imagery and the more primary perceptive and motor faculties in cognition (Damasio 1989). This agrees with some research on memory which suggest that memory for events is closely linked with motor memory, something which has also been documented in musical cognition (Mikumo 1994).

From research in auditory scene analysis, it seems reasonable to conclude that there also is a close cross-modal link between sound and action in musical imagery based on learning. We could look upon these sound/action associations as generalizable, and assume that there is a kind of "naive physics" cognition of sound/action association going on in most musical imagery, such as in the case of the mallet hitting the drum, an empty bottle tumbling down a staircase, an orchestra playing a tutti chord fortissimo, etc. In this connection, action can furthermore be seen as inherently hierarchical in the sense of having an executive intentional control with several more automated subordinate control routines (Holding 1989). Action can thus be seen as composite, consisting of an executive initiating impulse and a continuous trajectory in time-space, having in this sense both a discontinuous aspect and a continuous aspect. Being
chunked in terms of control, yet also continuous in terms of execution, action is then in fact a paradigmatic case of the relationship between the discontinuous and the continuous, shedding light on the above mentioned enigma of the continuous and discontinuous musical cognition.

3. Musical quantas and enactment

The emergence of relatively stable entities on the basis of flux substrates has been one of the main topics of phenomenological thought. In the work of Husserl, this is accounted for with a tripartite model of retentions, primal impressions and protentions, meaning that the continuous stream of impressions (primal impressions) are accumulated (in retentions), and that retentional images also will influence the perception of that which is to come (protentions). However, even though this is a continuous process, Husserl (inspired by Brentano) claimed it is intuitively evident that images of what has passed must emerge discontinuously, in a series of "nows", otherwise there would be no images at all, but only a continuous stream (Husserl 1980). A similar idea of discontinuous updating of images is proposed by Ricoeur with his lucid account of the need for "interrupting" the flux of experience in order to signify it (Ricoeur 1981), and interestingly, we find a related idea in the music theory of Xenakis with his terms of "in time" and "outside time" (Xenakis 1992). We could even see this as related to the concept of the window in digital signal processing, where the chunking of acoustic flux is the very condition for having spectral representations of sound. Although this is at the level of micro-resolution, the idea of the time window has also been applied at higher levels of cognitive modelling in music theory (Leman 1995). Even if we assume that cognitive processes are continuous in the sense of a continuous reception and transduction of signals from both the outside world and from within our bodies, it seems difficult to imagine that there is not also such a discontinuous chunking of sensory input in the cognitive apparatus. We thus have good reason to speak of musical quantas in the sense that discontinuous chunks are formed in our cognition on the basis of intermittently interrupting the continuous acoustic flux.

These musical quantas can then be associated with sound-producing actions, where the actions have a similar quantal aspect in the sense of being unitary as intentions and as images of trajectories. Associating musical quantas with action quantas is perhaps obvious (e.g. drum sound associated with trajectory of mallet hitting the drum), however taking this one step further, we can see that this is consistent with the motor theory of categoric perception (Stevens 1972, Harnd 1987), and is consistent with categorization in general, where common motor schemata seem to be a strong factor in the forming of prototypes (Rosch et al. 1976). What should be investigated further though, is the extent to which actions and/or other bodily schemata influence chunking and coherence in musical imagery. It would be interesting to see how various gestalt principles such as closure, good continuation, exclusive allocation, etc. relate to action.
This could in particular be informative with regards to rhythmic grouping, where for instance distinctions between alternate 3/4 and 6/8 groupings of a stream of eighth notes could be seen as a matter of two different action inclusions. Besides such topics of prototypicality and criteria for coherence, the paradigm of action could be a strategy to investigate the dynamic features of musical imagery as well, such as scalability (fast or slow replay of musical objects, highly compressed or highly expanded images of musical objects, variable resolution of musical objects, etc. by corresponding images of actions), as well as generativity or transferability of action schemata to new musical objects.

The mentioned main reasons for including imagined sound-producing actions in music theory (stimulus support in imagery and schematic ordering of acoustic signals into coherent musical objects), could be seen as integrating the idea of enactment in music theory. Conceptually, this may be difficult to accept, because music theory, like most western science, has a preference for making verbal or other symbolic representations. However, I believe there now is fairly good support for accepting the epistemological role of enactment (Varela, Thompson and Rosch 1991), and this could then open for further investigations of the relationship between sound and action, as has been done in the remarkable introspective study by David Sudnow on jazz improvisation (Sudnow 1978). There seems to be some material on enactment in cognition to be drawn from the field of visual imagery, in particular concerning the generation, maintenance and transformations of images (Kosslyn 1994), and there seems to be much material of interest in the field of research on motor behaviour (Holding 1989), and finally, we should also welcome more work dedicated to the kinematics of music making and musical expression (Todd 1995). However, as a conclusion so far, I believe it will be both necessary and fruitful to pursue this work on chunking in music theory by imagined sound-producing actions also by analogical representations, i.e. by research based on introspective observations of musical objects along the lines of Schaeffer, but in the future hopefully with the supplement of more refined graphical means for representing trajectories in timespace.

References
Cambridge, Mass.: The MIT Press.
I. Introduction
Despite the fact that polyphony is one of the most essential aspect of musical language, there are very few cognitive models which cover the perception of it; even in music theory. This is partly a consequence of tight relations between music theory and linguistic theory. But it is also a result of the fact that cognitive modelling is narrowed to linguistics as a main source when it comes to human expression.

In this paper, I want to show that music is a multidimensional form of human expression. By investigating overlapping structures in music as a borderline between melodic onedimensionality and polyphony, I want to demonstrate to what extent polyphony has to be regarded as a fundamental aspect of a musical grammar. I will then use the Generative Theory of Tonal Music (Lerdahl & Jackendoff 1983) as an example of a music theory which does not include the multidimensional aspects of music because it is regarded as a kind of deviation.

I will then present some borderlines in linguistics by focusing on the metaphor. This will show the need for alternatives in cognitive modelling in general. As one example of an alternative I will present the American philosopher Paul M. Churchland and his special version of a network-theory.

II. Music as multidimensional
Ravel's Bolero is an example of a linear movement which should be interesting in this connection. Very much of the musical information lays in the melodic line. The theme may easily be divided into two parts; each one of them containing eight bars, plus one-eighth note in the second part.
Ex. 1. Ravel: Bolero, Flute, bars 5-21

The long ending note and the pause in the first phrase make this division easy. It would be naturally to make a parallel division on a smaller level after the first four bars. That is not so easy because of a lack of symmetry. If each part should be divided into smaller phrases, the end of the phrase would represent a kind of understatement. This kind of understatement is present in the second bar of the theme.

Ex. 2. Flute bars 5-6, bar 6 is moderated at the end

The beginning of the next phrase, therefore, coincides with the end of the first phrase, and there is an overlap between the two phrases.

Ex. 3. The same kind of divisions with overlapping structures might be done in bars 7, 8, 9, and 10. According to the Generative Theory of Tonal Music, these structures represent a kind of deviation (Lerdahl & Jackendoff 1983). This piece of music might represent a good argument for this perspective. The form of the accompanying snare drums is very symmetrical with its two-bars structure.

Ex. 4. By comparing the flute with the drums it is demonstrated a sort of complementarity between the two voices. The regular form of the drums represents a kind of deep structure which generates the melodic voice as a counterpoint. By being a counterpoint to the drums, the melodic theme has to be syncopated and transformed in a way that generates its asymmetrical structure.

The relation between syncopations and counterpoint in general is commented by Heinrich Schenker. He says that syncopation is a result of how the voices are leaded, and syncopations therefore, are intimately connected to the counterpointal way of thinking in polyphony (Schenker 1978, pp.118ff.). According to Schenker, the polyphonic structure has to be regarded as a fundamental aspect of music. As a consequence, the syncopations, therefore, are subordinated to polyphony, but they represent also the same kind of fundamentality in music as polyphony does.

It seems to be a similar relation between overlapping structures and polyphony too. If we put the examples number two and three together on the same staff, they will form a naturally polyphonic structure with two voices.

Dr. art. Hroar Klempe: Polyphony as a challenge to cognitive modelling, 3rd ESCOM 1997
This represents an alternative notational form to the generative theory of tonal music, which would have supplied the melody with one added bar for creating a well-formed structure.

Example 5 illustrates that the deep structure of an overlapping event is polyphonic. On a surface level, therefore, one may say that the overlapping structure in melody represents a kind of implicit polyphony: two voices are been condensed into one. Whereas the notational form in example 5 underlines the aspects of polyphony, example 6 stresses the opposite, namely that overlapping structure has to be considered as an one-dimensional, melodic event.

There are many reasons for anchoring overlapping structures in polyphony. One is that there seems to be a very close relation between overlapping structures and intensificational parts in music. In the sonata form, there is a coincidence between overlapping structure and highpoint in the development. This technique is even more known from the baroque fugue as a stretto, which represents an intensification of the polyphony.

In other words: Overlapping structures have to be regarded as rooted in polyphony rather than in the melodic line because overlapping structures seems to be generated from an underlying polyphonic pattern. This means that polyphony, compared to melody, must be treated as a more fundamental aspect of musical language. It implies also that musical language represents a kind of multidimensionality as a fundamental part of it. This makes it different from language which must be regarded as representing a kind of one-dimensionality.

III. Schema-modelling as an onedimensional approach

The notion of wellformedness is strongly connected to a question of rationality. It is a question of avoiding ambiguity. This is central and goes back to Chomsky and transformational grammar. The deep-structure level in transformational grammar is connected to semantics. The solution of an ambiguous sentence lays in the fact that it is derived from two different deep structure sentences (Chomsky 1965/76, p. 22). Wellformedness on the level of deep structure thus, is a question of keeping rationality. The clearer meaning, the more wellformedness. On a surface level, that is on the level of performance, there might be a lot of ambiguous sentences. This means that two, or even more sentences might be condensed into one. On the level of performance there might appear a certain degree of multidimensionality. On a deep structure level, only one-dimensionality is accepted.

This corresponds very much to the way David E. Rumelhart defines a schema. He says that «a schema theory is basically a theory about knowledge» (Rumelhart 1980, p. 34), and he draws the lines back to Kant (ibid. p.33). He continues by stating that for each person, schemata represent the fundamentals in the process of comprehension. The schemata, therefore may be compared to theories. They are the way we interpret the world and therefore serve the function of
being each person’s unarticulated, private theory (ibid. p.37). The method for choosing a schema is the same as for choosing a theory, that is a deductive-nomological method in the sense that hypotheses are tested and rejected by observation.

Nevertheless, it is necessary to make distinctions between different kinds of schemata. Jean Matter Mandler distinguishes between what he has called 1. A story schema, 2. An event schema, and 3. A scene schema. The first specifies the general form and sequence of the events that occur in simple stories such as folktales and fables (Mandler 1979, p. 266). Event schema, on the other hand is defined as: «temporally organized representations of common sequences of events» (loc.cit.). Another word for event schema is «script». This term signalize the sequential aspect of an event schema. And it is quite concrete in the sense that it is a detailed description of an event, like, for instance, visiting a restaurant etc. The third form of a schema, the scene schema is defined as «a cognitive representation of what one expects to see when viewing (or entering) a scene» (ibid. p. 264). This is not sequential in the same way as an event schema. It is about the different elements that form a scene, like furniture, surroundings and so forth.

Mandler states that musical schemata share many of the characteristics of an event schema (Mandler 1984, p.86). They are both temporally organized, and they are also both ordered in accordance with a hierarchy which generate the sequence. He refers to some of the earlier researches made by Carol L. Krumhansl et al., where the hierarchy of stability within a set of musical tones forms a central point. Like in other event schemata, the hierarchy in music makes it possible to narrow the amount of possibilities in fullfilling a musical sequence.

Nevertheless, Mandler points out some differences that exists between music schemata and event schemata too. The differences are connected to the aspect of goals. An event schema is marked by different goals that are ordered in a hierarchy: «The first action was done in order to do the second» (ibid. p.82f). This must be understood as a kind of causality which represents a hallmark of an event schema.

What is a problem then, is that causality, in general, presupposes a sort of one-dimensionality. The rationality of causality is a result of a final cause, out of which everything else might be explained. And as far as the event schemata are defined by causality, they are also defined by the same sort of one-dimensionality. The reservation Mandler does in comparing music schemata to event schemata is interesting because it seems to be more grounded on intuition than on rationality. Music seems to contain at least one aspect which is not covered by the event schemata. And as far as Mandler defines event schemata in very close connection to what other theorists have called «scripts», one-dimensionality appears to be a salient feature of it.

Causality is one aspect that generate this feature. Verbal narrativity is another. The way Mandler and others expose the nature of script is by analysing the story in which the event is told. The stories are adapted for the purpose of demonstrating what a script is, and what is left is pure information, where noncontradictory formulations, which have clear denotations and are ordered in a sequentional pattern, form prevalent characteristics. Even the term «script» refers to language as the main source for defining what it is. The characteristics that are focused on here show that the notion of event schemata first of all are developed in considerations with language. And it is a traditional perspective on language which excludes the vertical aspects of multidimensionality. And this way of modelling the cognitive structures might even be inadequate for language.

Anyway, it is inadequate as a model for perception of music because it does not posit any hooks on which the aspect of musical polyphony might be hung.

IV. Alternative cognitive modelling

The need for alternatives in cognitive modelling is formulated by many. One of the main reasons lays in the fact that everyday language is not unambiguous. This is pointed out by several linguistics (Bernstein, Halliday etc.). And in contrast to Chomsky, Michael Halliday states that metaphor «is a feature of languages, not just antilanguages (sic.)» (Halliday 1978/94, p.177). George Lakoff and Mark Johnson state that metaphor is not only «a matter of words». On the contrary, metaphors are central in conceptualising the world: «they provide a means of understanding one kind of thing or experience in terms of another kind» (Johnson, M. and G. Lakoff 1982, p. 1). This is true for several events that are grasped by language through abstract concepts like «time, work, the mind and emotions, social and interpersonal relationships etc.» (loc. cit.). And as far as metaphor is defined as A is B (Danesi 1993, p. 123, Johnson and Lakoff op. cit, p.5), it implies a certain degree of linguistic ambiguity.
The American philosopher Paul Churchland is criticising traditional cognitive modelling in a perspective of theory of science. His point is that explanatory understanding happens so fast that there can not be time for applying a rule governed deductive-nomological model. (Churchland 1989, p.199.) Explanatory understanding, he says «is often achieved almost instantaneously, as when one understands at a glance why one end of the kitchen is filled with smoke: the toast is burning!» (Loc. cit.) It is the same in science. If one is well enough skilled in a certain field, the explanation appears in the same speed as if it was an everyday life event. Instead of criticising the deductive-nomological model from a logical point of view, he does it from a psychological point of view.

What he presents as an alternative is what he calls «parallel distributed processing (PDP) models of brain function» (ibid. p. 196). The process is described by several different input vectors that are activating several different «hidden units» in a network which produces an output. This is a network theory which is very different from a schema-theory: first of all because the schema is replaced by a synaptic connected neural network. This means that several different cognitive layers are working and cooperating at the same time as soon as they are activated by an input. Because very many hidden units are activated at the same time, there is a multiplicity which does not produce rationality and exactness. On the contrary, for the unique experiences it produces a certain degree of chaos. When the experience is repeated, it activates a certain synaptic pattern in the hidden units which becomes familiar after a while, and then seems to stabilise the situation. This produces a kind of output, but this does not have to be conceptualised or unambiguous. On the contrary it might be a very complex kind of expression, like a metaphor, a neologism or even a scream. The explanatory conclusions are reflecting the same kind of multidimensionality that rules the cerebral process. Despite the fact that the conclusions may have explanatory power, they do not necessary have to be precise or unambiguous.

V. The adequacy of alternative cognitive modelling in music

This model may correspond to music. As have been pointed out earlier, very much of the musical events are rooted in a musical polyphonic structure, even in melody. Music, therefore is not only an event that runs through time. There is a kind of simultaneity that may be compared to a scene. But the scene must be like the one depicted above, as when a kitchen is filled with smoke. That is a very complex scene, and very much is happening at the same time.

On the other hand experiments show that attention cannot be given to more than one melodic line at the same time (Sloboda 1985/89, p. 169ff.), and in a counterpuntal movement the other lines are only registrated as a kind of harmonic support to the melodic line. One may ask then, to what extent a network-modell may represent an adequate explanation of the way music is perceived?

The immediate reaction to this question is that there are a lot of polyphonic music, and one may ask: why should any write polyphon music if it was not perceived as polyphonic? In other words: this music must include some qualities that must be registrated by the listener, not necessarily as different melodic lines with the same clearness as one melodic line, but nevertheless, registrated as different melodic lines.

As has been pointed out earlier in this paper, clearness in itself is strongly connected to one-dimensionality and unambiguity. When Mozart, then, makes fun out of music in his little diverting piece of music, Ein musikalischer Spaß (KV 522), it is exactly because too much things are happening at the same time: Too many styles are mingled together in a very unfullfilled mixture. In the opening part of the first movement, this effect is also created by an overlapping structure in the first line.

Ex. 7. Mozart, Ein musikalischer Spaß, KV 522

Dr. art. Hroar Klempe: Polyphony as a challenge to cognitive modelling, 3rd ESCOM 1997
Humour is unpredictable and ambiguous. But unpredictability and ambiguity are qualities which are not only reserved for humour in music. In the Chorale St Antonio, which is used by Brahms' "Variationen über ein Thema von Joseph Haydn" one may find a similar sort of overlapping structure, which may be divided in this manner:

Ex. 8. Brahms: «Variationen über ein Thema von Joseph Haydn»

In other words: Polyphonic structure in music seems to avoid exactness. Instead it creates a kind of vagueness in the sense that attention has to be switched over to another perspective as soon as a new voice is presented. Two different voices do not need to be registrated, but as far as memory is an accompanying part of attention, polyphony is creating a kind of multi-dimensionality in the mind. According to a network theory, the human mind may be modelled in a way that includes this kind of multi-dimensionality. And as far as polyphony is a fundamental aspect of the musical language, it seems to be a human means of expression that do not only correspond to this multidimensionality, but also evokes it.

References
Mandler, Jean M.: «Categorical and Schematic Organization in Memory» in Puff 1979
Schenker, Heinrich: Free Composition, , Longman New York 1978
Spiro, Bruce and Brewer (ed.): Theoretical Issues in Reading, Hillsdale New Jersey 1980

Dr. art. Hroar Klempe: Polyphony as a challenge to cognitive modelling, 3rd ESCOM 1997
Music and Bodily Movement, Dance
How to assess schoolchildren's performance in dance?

Gertrud Ericson, Department of psychology, Uppsala University, Uppsala

Introduction

Dance as an art subject has been scarce in Swedish schools and research about its influence on children's development is almost totally lacking. The Swedish Ministry of Education and Cultural Affairs launched a "Culture in school" initiative in 1985. A multidisciplinary research project "Dance in primary school" got funding to start systematically to describe and analyze dance education in a natural setting. Prior evaluation studies of a municipal dance project in Enköping provided an empirical basis on which to pursue. A psychological perspective of dance as a multiperceptual stimulus situation was used as well as a perspective of "multiple intelligences" (Gardner, 1985) in outlining a tentative theoretical framework, and an expert panel of dance teachers contributed in order to identify aspects and variables to be observed. This paper is based on the doctoral dissertation "Assessment of Schoolchildren's Performance in Dance" (Ericson, 1996).

Method

Data were collected primarily through video recordings of ordinary dance lessons given once a week by a professional dance teacher and a skilled pianist in primary school classes 1985-88. The main data collection was made in 1986/87 in 16 first grade classes at the beginning and at the end of the school year when also tests of gross motor ability and self-esteem were administered.

Nine qualified observers worked out variables and scales for observation: a psychologist and academic teacher of psychology, a student of art and dance, an experienced classroom teacher and six students of psychology. Each of the selected variables was assessed on a 7-point scale, where 7 indicated the highest, 4 the average and 1 the lowest level of performance. Dance sequences that recurred throughout the school year were chosen for observation.

Interviews were made with two dance teachers, representing the living dance tradition from which this dance education was developed. Teachers were interviewed and parents got questionnaires. Interrater reliabilities were estimated by use of analysis of variance (ANOVA). Ratings were subjected to significance tests and further to exploratory factor analyses.

Results

The basic dance course

The dance education studied originated from a Russian classical ballet dancer, Vera Alexandrova, and was further developed by the dance teacher Eva Dahlgren. Certain elements originating from Alexandrova and Dahlgren are still traceable in the basic dance course: The children's dance has to come from within themselves and should have a meaning for every individual. When the dance is performed it will be shared through a sense of community, including everyone in the group. The dance teacher has to be emotionally present to establish a bond between the teacher and the children. They become absorbed in the special atmosphere, which ensues through a kind of suggestion.

Later, the dance teacher Cecilia Dahlgren developed a pedagogical method "shock pedagogy," characterized by short, rapidly changing exercises, high tempo, and generous encouragement with comments like "good," "fine," "delightful" throughout the lessons. A dynamic structure in the lesson ensues. A bodily "ABC" of dance is learned by the children to be used in
improvisations and performances for parents and schoolmates. The dance teacher shows a
movement or pattern of movements and may use her voice to create a certain atmosphere, or
verbally to arouse associations to objects, situations, places, experiences and emotions by use
of images and symbols. A dance lesson consists of 15-30 different sequences, covering a wide
range of contrasts and variations. Sequences that recurred throughout the school year were:

- basic movements; exercises for strength, coordination and flexibility; basic steps;
  formations and exercises for use of the space
- choreographed dances
- improvisations and the children's own dances.

The basic dance exercises skipping, sideways galloping, and walk and clap recurred every
lesson and were therefore chosen for observation.

A structural framework for the dance education was provided by the dance teacher together with
the music that embodied and represented a rhythmic basis for both the whole dance lesson and
its separate dance sequences. A pianist accompanied the dance, playing music that should
represent the same character of movement as the one presented by the dance teacher in her
movements.

Aspects and concepts for assessment

In most school subjects, instructions about what to do, and how, are mainly given verbally and
finished before the exercises get started, unlike in this dance education, where the pupils are
instructed and supported both by dynamic visual (the dance teacher) and auditory (the music)
models during exercises. The tasks to be performed in dance are mostly common to all pupils,
which means that not only is there one visual model (the dance teacher) but rather all the class
mates act as models to the actual movement, pattern of movement and its quality.

This situation of multiperceptual stimulation could be analyzed in terms of continually changing
visual-auditory and, to a lesser degree, kinesthetic perceptions of the dance teacher, that give
immediate bodily impulses of movement to the pupils. They learn how to transform the
primarily visually perceived impulses into individual, similar movement patterns. Initially, the
pupils learn mainly through imitating the dance teacher's movements and rhythmic patterns,
repeating and successively getting the movements or patterns of movement automatized. Later,
the pupils will also use the impulses to move as incentives to create their own individual or
group dance improvisations. In addition, the corresponding music played on the piano acts both
as an auditory and a bodily-kinesthetic stimulus, perceived by the pupils as sounds and
vibrations in the room, which are amplified by the movements of the dance teacher and the
children (e.g., stamping of feet and hand clapping). The pupils are surrounded and supported
by a rhythmic structure in almost all exercises.

Moreover, in certain exercises the pupils stimulate each other bodily-kinesthetically as they hold
each other's hands or waists, when, for example, dancing all together in a ring or in pairs. At
the same time the pupils would be influenced by visual-auditory perceptions of their peers
moving rhythmically, all in a ring or in pairs.

In all, this unusual multiperceptual stimulus situation can be expected to have specific effects on
the pupils. In addition to influencing the motor behavior of the children, their music behavior
could also be expected to be affected. Further, the continuous flow of multiperceptual impulses
in exercises holds and directs the physical behavior of the pupils. Initially, the moving dance
teacher is the focus of their attention. During the school year, inward personal feelings and
ideas of each child come more into focus. In the expert group, the movement of the child,
rhythmically related to the music, was considered central for assessment. After discussions with
the expert panel, due consideration taken also to areas of schoolchildren's performance in dance
where influences could be observed as indicated by parents and teachers, it was concluded that
schoolchildren's performance in these dance lessons may be adequately studied with regard to
four fundamental aspects: motor, musical, personal-aesthetic, and social. Using the circle as
In each aspect, a number of variables were identified and operationalized in terms of 7-grade rating scales. The variables included in the motor aspect were coordination, gesture, posture, effort, and body extension; in the musical aspect, they were rhythm and pulse, movement character, and phrasing; in the personal aspect, involvement, form, gesture, attention, willingness, perseverance, joy; and in the social aspect, integration and social atmosphere. A total impression of the performance of each child was assessed by the variable wholesomeness in the motor as well as the music aspect.

Analyses of judgments
Judgments of the variables in each aspect were made by two independent judges. The interrater reliability was good or satisfactory for practically all variables ($r_k > .60$ and/or $MS_{res} < 1.00$).

There were significant increases across dance classes from pre to postobservation ratings of group performance in almost all personal-aesthetic and social variables. Also in the music variables there were mostly higher post than preobservation ratings, although they were usually not statistically significant.

A number of organizational conditions were considered as potentially important for appropriate teaching in dance (e.g., professional dance teacher and pianist, suitable room). The 16 classes selected for observation were divided into two groups based on median-split, one group having favorable and another group having unfavorable conditions. There were no significant differences in gross motor ability or in self-esteem between the groups. However, classes having favorable teaching conditions showed larger increases in most variables from pre to postobservation ratings than classes having unfavorable conditions.

An exploratory factor analysis on personal-aesthetic and social variables combined indicated four meaningful factors interpreted as a "gesture-form" factor, a "social atmosphere" factor, a "joy" factor, and an "attention-integration" factor.

In most judgments of individual performance in the motor variables postobservation ratings were significantly higher than preobservation ratings. An initial difference between the genders in favor of the girls diminished or disappeared in postobservation ratings.
For girls, a significant development was found in every motor variable in *skipping* (except *posture*). For boys, a significant development was found in every variable in *sideways galloping* (except *effort*). Also, ratings of the music variables (*rhythm and pulse, movement character*) showed a similar gender-differentiated pattern: boys significantly developed in the pair-dance sequence *sideways galloping*, whereas girls significantly developed in the individual dance sequence *skipping*.

An exploratory factor analysis on motor and music variables combined resulted in three distinct factors that could be meaningfully interpreted: a "music/rhythm" factor, a "motor-energy" factor, and a "motor-posture" factor.

**Discussion**

**Data collection in a natural setting**

Data collection in a natural setting inevitably encounters certain problems regarding control of external conditions, possibly influencing the variables studied.

Controlled laboratory settings have been used in developmental psychology whenever possible and regarded necessary in order to obtain reliable answers. However, in recent years the transfer of results from a formal setting to a natural setting has been called into question as being less unequivocal than was previously thought. Artificiality in laboratory settings or formal settings in general may yield measurements with uncertain relation to real-life behavior (Gardner, 1993; Miller, 1987).

Data collection in natural situations with visible operator and technical equipment might influence the observed flow of events and behaviors. However, the video recordings of dance lessons disturbed neither pupils nor teachers, as evidenced by teachers when the recordings took place and also later when they watched the video recordings. Rather than noticing any negative effects, the teachers even expressed a feeling of authenticity while watching the recordings.

**Reliability**

A general problem in this study concerns measuring capacity through achievement, that is, using observations of behaviorally defined variables that are theoretically connected to underlying competencies.

Some professional knowledge about the competence to be judged, as well as experience of working with schoolchildren, seemed important for proper use of the scales. When these criteria were met, satisfactory interrater reliabilities were attained in most cases. Still, further investigations using other judges and children should be conducted before any safe conclusions about generalization can be reached. It seems reasonable that the variables and scales could be reliably used by the present judges for assessment of other children. Probably, other judges could attain satisfactory reliability if given proper training.

That high reliability indices were reached also supports the conclusion that meaningful aspects, variables, and scales have been established. Moreover, the fact that there were significant differences between pre and post assessment in most variables, although still difficult to interpret in terms of cause and effect, points in the same direction.

**Aspects for observation**

Achievements in dance were observed through four basic aspects: motor, musical, personal-aesthetic, and social. These were also mainly in accordance with the perspective of multiple intelligences. The content of the aspects was identified and made explicit by defining a number
of variables to be used by observers utilizing 7-step rating scales. The question is whether the results obtained support this division or if the aspects may be revised. The results indicate that the division into four aspects (Figure 1) should be replaced by divisions indicated by the factor analyses and tentatively visualized as in Figure 2. Since there were two different factor analyses, the areas are comparable only within their respective half-circles (the upper half-circle for motor and music aspects, the lower half-circle for personal-aesthetic and social aspects), not across the whole circle.

The music/rhythm part is relatively unchanged, constituted as before by rhythm and pulse and movement character. The motor aspect is split up into two parts. The motor-energy part concerns the amount of body effort and extension, which has to be adapted to the music structure in dance sequences. The motor-posture part, including coordination and gesture, concerns how body parts are placed and work together. The resulting motor patterns are visible in body contours, showing movement patterns or forms. It is therefore placed adjacent to gesture-form (in the personal-aesthetic aspect) which includes individually interpreted forms as well as group shaping of different spatial patterns in the room. Attention/integration and joy, as being more of personal characteristics than of skills, therefore appear adjacent to each other. Willingness is included in attention/integration and, with a minor loading, in social atmosphere. Thus the social atmosphere area borders on the other side of the attention/integration area.

![Figure 2. The aspects of assessment revised.](image)

The model in Figure 2 is, of course, still highly tentative and incomplete. Its validity has to be checked in further investigations with other samples of children, teachers, dance, and music.

**Practical implication**

Dance was neglected in Swedish schools when our research project started. Since then, dance has become more frequent in schools. Only three municipalities had dance in school before the dance education here studied got started. At present, about 60 municipalities have dance in school in one form or another.
The same type of video recordings as used in our research project could be used in basic teacher training, especially for dance and music teachers and teachers of other aesthetic subjects, but also for primary school teachers. Since the variables have been related to an overall theoretical perspective on children's capacities, the defined variables would provide an attainable vocabulary both for examined teachers and teachers in training. Further, the theoretical perspective would facilitate a pedagogical discussion about relations between goals for lessons and observed achievements. Observations could also focus on type of exercise, instruction, and subsequent behavior of the class, of boys and girls, respectively, learning styles, etc.

Future research

While getting into a new and interdisciplinary research area it was necessary first to obtain a thorough description of the phenomena (children's dance behavior) under investigation in order to get a basis for further theoretical and methodological elaboration. Therefore, the variables and scales presented here should be used in replication studies with other classes, dance teachers, music, and observers. This work will, of course, require many years of continued work.

Although it basically does not depend on research results to legitimize its place in school, dance education does have to be described and evaluated, like other activities, in order to secure its proper place in the debate about learning and development of schoolchildren.

References


Expression in Dance Performances
Ingrid Lagerlöf and Marie Djerf
Department of Psychology, Uppsala University, Sweden

In the artform such as dance the intention to be expressive is built into the activity. That counts for educational as well as professional dance. There is the intention to represent or express moods, feelings or concepts beyond the sheer movement of the dancer's body. Yet, we don't know much about how this intention is transformed into movements or apprehended by the spectator. That also raises the question whether the perception of dance is something subjective "in the eye of the beholder" or if there is a general agreement on the content. These questions were explored in two different studies. Study 1 will be outlined in more detail, while Study 2 will be summarized in a preliminary presentation.

Study 1

There has been an intense debate about the role of expressiveness in educational dance. Expressiveness as self-expression has been much criticized as serving emotional development and therapeutic goals, while aesthetic qualities has been dismissed (Haynes, 1987; Smith-Autard, 1994).

The standpoint of this presentation is that expressiveness is still an essential part of dance education and that it belongs to the dimension of aesthetic qualities, not as a mean of self expression but as a symbolic formulation of feelings, moods and concepts in accordance with the theory of S. Langer. This is illustrated in the following glimpse at a dance class. A theme of a improvisation was "walking on moon" which corresponded to the atmosphere of breathlessness, stillness and excitement. Student is encouraged to reflect these feelings in her/his personal way of movement, although movement quality was given by teacher. That presupposes that student constructs a cognitive concept (of this atmosphere) as well as a personal interpretation of this particular mood. Mapping of feelings is understood in coherence with symbolic order as well as it is related to a personal experience of these feelings. In this sense the performance is viewed as an expressive symbol and consequently it should be intersubjectively observable.

In order to locate the expressive symbol it was distinguished from two adjacent dimensions of dance performance; a) formal symbol and b) movement quality

a) A formal symbol refers to a physical object. The symbol is a clear referent to obvious characteristics of the object, like size, spatial dimensions or movements

b) Movement quality is without any symbolic content. That is, it relates to the dancer's ability to combine elements of time, energy and space into movements.

Expressive and formal performance are different regarding dancer's intention with expression, whereas movement quality is a technical component fundamental to all movements without any symbolic meaning of its own.

The emotional and expressive involvement in dance also concerns the traditional female and male manner of movement. There is an overall tendency to assign emotional involvement to the female manner. Gender typing attitudes also seem to incorporate body language. Tender and lyrical movements are defined as female manner, whereas large, determined and forceful movements would be labeled as male manner. And it is learned from dance education that it could be provoking to dance against tradition.
The overall aim of the study was to distinguish expressive quality from adjacent dimensions of dance movement. This is done by operational distinctions between the categories of expressive, formal and movement qualities.

**Method**

One hundred and nine seven-year-old children 50 girls and 59 boys, were video-taped, while improvising on different themes. Recordings were made in natural school setting on 13 regular dance classes once a week. Eight recordings were made in the beginning and five recordings were made at the end of the school year.

A videocamera was placed in one corner of the room and danceclass was videotaped. All sections of improvisations were then selected and copied on to videotape for analysis. Units for judgements ranged between 5-35 seconds.

Children were instructed by dance teacher to improvise on different themes like "robots", or "walking on moon". These themes were also made to correspond to different state of feelings, emphasized by piano music. "Robots" appeal to strength, energy and aggressiveness and "walking on moon" appeals to breathlessness, stillness and excitement. The purpose was to direct childrens attention to a feeling-form.

Movements were instructed by the danceteacher and all children performed the same sequence of movement, i.e. sequence of movements was a constant variable.

Children would differ on motivation and ability for the task. Therefore it was assumed that children could either depict a pure "formal" performance; the objective characteristics of the object (jerky movements) whereas some would also enter into expression of feelings. Moreover, sterotyped (conventional manners) and divergent performances were used as control variables. "Movement quality" was an additional distinction, i.e., technical ability, based on combinations of time-space-energy, which is fundamental to all dance movements. Judges were instructed to categorize performances according to following definitions.

- individual's performance who transmitted an emotional presence and meaning to spectator was categorized as expressive performance

- individual's performance who did not transmit any emotional presence, but is concentrated and correctly reproduces the characteristics of a realistic object was categorized as formal performance

- individual's performance who carried out a stereotyped movement according to social demand and also showed certain lack of concentration, for example, wanting to be in physical contact with other students, was categorized as stereotyped performance

- individual's performance who simply did not follow direction from danceteacher was categorized as divergent performance

Three independent qualified and three unqualified judges categorized children according to instructions above.

Two qualified judges rated movement quality for 67 individuals on a seven point scale.
Results

Interrater reliability for expressive and formal categories alone was Kappa=. 83 and for all four categories interrater reliability was .81 among qualified judges. Unqualified judges reached lower interrater reliability.

Table 1. Interrater reliability as Kappa (K).

<table>
<thead>
<tr>
<th>Performances</th>
<th>Qualified judges</th>
<th>Unqualified judges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expressive/formal performances</td>
<td>0.83</td>
<td>0.50</td>
</tr>
<tr>
<td>Expressive/formal/stereotyped/ divergent performances</td>
<td>0.81</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Table 2. Interrater reliability as Pearsons correlation (r) for movement quality

<table>
<thead>
<tr>
<th>Interrater reliability</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Movement quality (n=67)</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Correlation between expressive performance and movement quality was \( r_{bis} = -0.27 \).

About half of all movements were categorized as expressive performances, see figure 1. More boys than girls made an expressive performance, and far more girls than boys made a formal performance, see figure 2. There was a significant gender difference for stereotyped and formal performances (\( \chi^2 = 10.0, p = .02 \)).

![Figure 1. Distribution of performances in percent, n=109.](image)
Discussion

For qualified judges interrater reliability was high when distinguishing expressive performance in dance education as well as for movement quality. Results show that expressive performance is perceptible within very short time units, ranging from 5-35 seconds. Interrater reliability was not satisfactory for unqualified judges.

Correlation between expressive performance and movement quality was low, which means that movement quality by no simple means determines the perception of expressive performance. Thus, expressive performance is not the same as high technical skill. This also indicates that technical training is not sufficient in order to reach an expressive performance.

We do not know much about development of dance competence. However, according to these results it seems that seven-year old children are able to symbolize complex feelings in dance. Therefore it is possible to work with the symbolic expression of feelings much in line with what Abbs (1989) describes as the reciprocal interaction between impulse and symbolic order at an early age.

Activities that emphasize emotional involvement are traditionally defined as belonging to a female domain. Therefore it is interesting to note that expressive performance is not less but rather more prevalent among boys in this study. Moreover it was mostly girls who made a formal performance.

Study 2

Aim of Study 2 was to investigate how professional dancers' intention is transformed into movements when expressing different feelings and the concepts of femininity and masculinity.

Method

Six dancers, 3 dance students at the Swedish College of Dance (all female) and 3 professional dancers (2 females and 1 male) participated in the study.

Dancers were instructed to improvise on concepts of feelings like anger, grief, joy, surprise, fear and also the concepts of femininity and masculinity, which they then performed individually to the accompaniment of a pianist. Improvisations and performan-
ces were videorecorded. Dancers watched the recordings and each student was then interviewed about what strategies they used to reach their final performance. Dancer also rated their degree of emotional presence at different performances. The videotaped performances were also presented to an unexperienced group of subjects, who described their experiences of the performances by free associations and ratings.

Results

Preliminary results show that improvisations were based on three different strategies for transforming ideas into movements; 1) to depict a concept where specific characteristics were picked out and evolved into movements, 2) to start with a movement quality, for instance "floating", which was elaborated to a movement form, 3) to start from "lived body experience"; an embodiment of mind and body, i.e., mind was not single source or mover for movement. Dancer turned to a personal embodied experience of the intended concept, for shaping movement.

There were differences in focus, symmetry and effort in movements when perceivers labeled performances as either feminine or masculine. Focus of "masculinity" was directly confronting perceiver whereas focus of "femininity" was turned away. "Masculinity" was composed of symmetrical movements with strong effort, but "femininity" had more of nonsymmetrical movements and relaxed effort. Performances of femininity had more complex and changing movements, in contrast to more clear-cut and larger movements for masculinity. There were also differences in using of space and postural stance at beginning of performances.

Spectators' ratings showed high agreement. Dancers' strategy of "lived body experience" shows a high correlation to spectators' experience of dance expression. Moreover dancers' own ratings of emotional presence at performances correlated positively with spectators' ratings.

References


COMPARING RUNNERS' DECELERATIONS AND FINAL RITARDS

Anders Friberg and Johan Sundberg

Royal Institute of Technology
Speech, Music and Hearing
S-100 44 Stockholm, Sweden
email: andersf@speech.kth.se, pjohan@speech.kth.se

INTRODUCTION

Music and motion are often assumed to be closely related as suggested, although the exact basis for this relation is unclear. The present experiment compares quantitative measurements of final ritards in musical performance and of the stopping of running.

The characteristics of the final ritardando in music performances were analyzed by Sundberg and Verrillo (1980), who measured also computed an average tempo profile. Later similar curves were observed for the tempo changes used for phrase marking in music performances (see e.g., Todd, 1985; Repp, 1992). Kronman & Sundberg (1987) found that the average final ritard profile could be rather accurately approximated by a square root function of score position. They also observed that the same function could describe the decrease of foot-step frequency of a stopping runner, assuming a constant body deceleration. Todd (1995) further developed such analogies between music and motion and extended the Kronman & Sundberg model for tempo change to phrasing.

The exact form of the ritard seems crucial. In particular, a smooth onset of the ritard and the impression that the last tone sounds as being the last in the piece seem fragile ritard properties.

Both Kronman & Sundberg and Todd derived their curves on the basis of theoretical considerations only, and no data were collected from footstep frequency patterns during runners' deceleration. The main aim of the present investigation was to collect such data and compare them with tempo changes during final ritards.

EXPERIMENT

Data on runners' decelerations were gathered in an experiment where two male and two female professional dancers were used as subjects (for a detailed account of this experiment, see Friberg & Sundberg, 1997). The subjects' task was to run at predetermined initial step frequencies along a marked straight line and to stop their running at a marked point. Two step frequencies were selected with interstep intervals of 250 ms and 340 ms. Five repetitions were recorded in all conditions. Audio and accelerometer signals reflecting the steps were recorded on a DAT PCM machine (TEAC). The running was also filmed by a video camera. Step length was determined from the video recording, and the foot-step durations from the audio and accelerometer recordings.
RESULTS

Rating experiment

All video recordings were first assessed in a pretest with regard to overall aesthetical quality by an experienced teacher of eurhythmics at the College of Music, Stockholm. The selection of stimuli for the main test was based on the results of these evaluations. A total of 16 decelerations were then selected taking into consideration variation with respect to step frequency, subject, pre-test rating values, and deceleration distance. These stimuli were copied onto a video tape, where each deceleration appeared 3 times. The tape was presented to 6 professional experts in choreography and/or eurhythmics who rated the overall aesthetic quality of the decelerations. The subjects were asked to rate overall aesthetic quality, taking into consideration wellformedness and balance. Each of the 48 tokens corresponded to a 10 cm line on an answer sheet where the left end represented “Very unaesthetic” and the right end “Very aesthetic”.

Table 1 summarizes the test result in terms of ratings averaged across replications and subjects and the corresponding standard deviations. The table shows that the subjects did not clearly prefer neither male or female dancers, nor any particular values of step frequency or deceleration distance condition. The four decelerations which received the lowest mean ratings were all produced by one male dancer, M2. Most of these exhibited irregularities in step frequency patterns. In the subsequent processing of the deceleration data we discarded these four tokens and analyzed only the remaining 12.

Table 1. Aesthetical quality of decelerations according to the evaluation test with experts. The data represent means and standard deviations across repetitions and subjects for the 16 decelerations selected.

<table>
<thead>
<tr>
<th>Deceleration number</th>
<th>Runner</th>
<th>Step frequency (Hz)</th>
<th>Deceleration distance condition (m)</th>
<th>Deceleration distance condition (cm)</th>
<th>Mean Rating (cm)</th>
<th>SD</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>M1</td>
<td>2.9</td>
<td>7</td>
<td>6.3</td>
<td>2.5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>M1</td>
<td>2.9</td>
<td>4</td>
<td>6.3</td>
<td>2.2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>M1</td>
<td>4</td>
<td>7</td>
<td>6.2</td>
<td>2.5</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>F2</td>
<td>2.9</td>
<td>F</td>
<td>5.9</td>
<td>2.5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>F2</td>
<td>2.9</td>
<td>F</td>
<td>5.8</td>
<td>2.5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>F2</td>
<td>4</td>
<td>F</td>
<td>5.8</td>
<td>1.9</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>F2</td>
<td>4</td>
<td>F</td>
<td>5.6</td>
<td>2.7</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>F1</td>
<td>2.9</td>
<td>F</td>
<td>5.3</td>
<td>2.4</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>F1</td>
<td>2.9</td>
<td>F</td>
<td>5</td>
<td>2.3</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>F2</td>
<td>4</td>
<td>4</td>
<td>4.8</td>
<td>2.1</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>F1</td>
<td>4</td>
<td>F</td>
<td>4.4</td>
<td>2.5</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>F1</td>
<td>4</td>
<td>F</td>
<td>4.2</td>
<td>2.3</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>M2</td>
<td>2.9</td>
<td>7</td>
<td>3.5</td>
<td>2.3</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>M2</td>
<td>4</td>
<td>7</td>
<td>2.7</td>
<td>1.8</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>M2</td>
<td>2.9</td>
<td>4</td>
<td>2.6</td>
<td>2</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>M2</td>
<td>4</td>
<td>4</td>
<td>1.5</td>
<td>1.8</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

Deceleration patterns

The step frequency patterns varied widely, some decreasing slightly, some remaining basically constant and some even increasing toward the end of the deceleration, thus
apparently negating the possibility that the change in step frequency serves as model for the
tempo change in final ritards. The step length, by contrast, decreased gradually toward the end
of the deceleration in all cases. Thus, the slowing down was achieved mainly by decreasing
the step length. When we previously calculated the idealized pattern for the deceleration of
running which almost perfectly matched the final ritard curve, a constant step length was
assumed (Kronman & Sundberg, 1987). The present results indicates that this was an
unrealistic assumption. The decrease of velocity showed a systematic behaviour was similar to
the mean ritard tempo curve, thus corroborating Todd’s idea of a parallelism between tempo
and velocity (Todd, 1995). Figure 1 shows the normalized mean velocity for the 12
decelerations which received the highest ranks (mean ratings > 4.2) together with the mean
tempo curve for the final ritard (Sundberg & Verrillo, 1980).

![Figure 1. Normalized mean velocity for the 12 highest ranked decelerations (solid line) and
mean ritard tempo (Sundberg & Verrillo, 1980). The abscissa represents normalized time.]

The almost perfect match illustrated in the graph was obtained by scaling the abscissa of the
velocity curve by a factor of 1.13. The use of this scaling is motivated by the fact that
obviously, the ritard curve can never reach a normalized tempo value of zero.

According to Kronman & Sundberg (1987, see also Todd, 1992), the deceleration force is
constant during the deceleration of running. This implies that velocity is a linear function of
time. As seen in Figure 1 this was not the case. However, the square of the velocity, i. e., the
kinetic energy, is close to a linear function of time, see Figure 2. This means that the power
used for the deceleration was approximately constant throughout the entire deceleration.

---

1 In a previous preliminary report on these results (Sundberg & Friberg, 1996) we erroneously used
"Normalized Score Position" rather than "Normalized Time" as the abscissa.
DISCUSSION

An important question is to what extent the experimental conditions in the recording session were specific to the velocity curves obtained. Obviously, we would have obtained quite different deceleration curves if the dancers had been running in a slippery surface, such as ice. The recordings were made in a gymnasium, where, presumably, friction was optimal and thus should have caused no abnormality in the deceleration data. Furthermore, in the rating experiment no preference was observed, neither for step frequency, nor for deceleration distance. Thus, our mean velocity curve can be assumed to be reasonably representative for the stopping of running.

It could be argued that dancers are prejudiced subjects in an experiment with the present aim, being prone to adopt the typical patterns of classical choreographic art. On the other hand, none of the subjects showed any tendency to apply any special locomotion patterns during the experiment. It could also be argued that professional dancers are untypical subjects as they have been specially trained to move in synchrony with music. However, professional dancers share with almost anyone the experience of moving to music, since dancing is very commonly enjoyed form of social life.

To rate the aestetical quality of a runner’s deceleration implies the task of giving a global assessment of a multidimensional stimulus. In our case this would have contributed to the comparatively high standard deviations of the ratings. The judges may have attributed different degrees of importance to various aspects of a given deceleration. Still, the rating experiment seemed to serve its main purpose, viz., to identify tokens which appeared unacceptable to expert judges.

The deceleration power was found to be approximately constant throughout the entire deceleration process. Thus, the dancers’ velocity patterns appeared to be well-planned, which should enable a spectator to predict the point of final stop. The similarity between the velocity curves and the tempo curves for the ritard supports the hypothesis that the ritard is an allusion.
to the stopping of running. Predictability of the instant at which the final tone appears in a piece of music is important not only to the players but probably also to the listeners. It is possible that the time of appearance of the final tone becomes predictable in music listening because of the similarity between the final ritard and a runner's deceleration curve.

If the above reasoning is correct we should expect a difference in tempo curves for a local rallentando inserted for the purpose of phrase marking and/or emphasis in general; an important task of a phrase marking rallentando is to inform the listener about the fact that the piece is going to continue. It would be worthwhile theme for future research to compare tempo curves for final ritards with rallentandos marking phrase boundaries. Likewise, comparisons between hand movement patterns, such as during conducting, and local tempo change patterns might offer interesting insights into the similarities between locomotion and music.

CONCLUSION

The mean instant velocity during deceleration can be accurately described by square root function of time. This implies that the runner is supplying a constant amount of deceleration power throughout the entire deceleration process. Such a deceleration curve is strikingly similar to a curve showing tempo versus time in final ritards of musical performances. This suggests that the final ritard is an allusion to a type of termination of locomotion.

ACKNOWLEDGEMENTS

Ulla Hellqvist, the Royal College of Music, Stockholm is gratefully acknowledged for performing the initial evaluation of the decelerations. The participation of the dancers and subjects is also gratefully acknowledged. Sofia Dahl, Jenny Ivarsson and Bo Lantz, assisted in the analysis of the video data. This work was supported by a grant from Swedish National Council for Research in the Humanities and Social Sciences.

REFERENCES

Music Cognition: General
Introduction: Cognitive Music Theory

Over the past 15 years or so, significant progress has been made in bringing the psychology of music and music theory closer together. Lerdahl and Jackendoff’s *Generative Theory of Tonal Music* is perhaps the most striking example of this, but Narmour’s work on melody (Narmour, 1990), Butler’s research on tonal structure (Butler, 1992), Meyer’s writings on a variety of issues over a considerably longer period (e.g. Meyer 1956, 1973), and from the ‘other side’ of this divide, as it were, Krumhansl’s research on pitch structure (Krumhansl, 1990), and much more besides, are all symptomatic of a desire that music theory should be relevant to listening, and that work in music perception should engage with the preoccupations of musicians and music theorists. There still remains, however, something of a yawning gulf between psychological accounts of our response to music, and those offered within aesthetics and critical theory - though there is arguably a greater degree of potential common ground between these two domains than between the psychology of music and music theory. There have been a number of responses to this: one has been mutual suspicion and rejection, with the music perception community dismissing aesthetics and critical theory as hopelessly rarified and unrealistic, preoccupied with the elaboration of its own hermetic discourse and unrelated to the facts of listening; and the aesthetic and critical community rejecting the simplistic and reductionist character of the psychology of music, and its lack of relevance to the sophisticated sensitivities of engaged and enculturated listeners: “Music recognizes no natural law; therefore, all psychology of music is questionable.” (Adorno, 1973: 32). Another response from the music perception community has been the ‘not yet’ approach: before we can deal with the highest levels of musical response we need to sort out some of the basic processes. This is essentially the position of Lerdahl and Jackendoff, for instance, when they write: “To approach any of the subtleties of musical affect, we assume, requires a better understanding of musical structure. In restricting ourselves to structural considerations, we do not mean to deny the importance of affect in one’s experience of music. Rather we hope to provide a steppingstone toward a more interesting account of affect than can at present be envisioned.” (Lerdahl & Jackendoff 1983: 8). And from outside the music perception community, there have been attempts to integrate, or graft on, bits of cognitive theory with philosophical aesthetics (e.g. DeBellis, 1995). None of these can really be regarded as having made much progress in bridging the gap - a gap which admittedly is regarded as inherently unbridgeable by some. The position I adopt in this paper is that the failure to make any progress can be attributed to the assumption that the cultural, aesthetic or ideological value or content of music is more remote than its basic perceptual properties. It is that assumption that I will challenge, along with the approach to perception on which it is based.

With a few exceptions the overwhelming majority of work in the music perception literature can be crudely classified as adopting a cognitive approach. Simplifying enormously, a fundamental principle in this approach is that perception has a sequential character, starting with basic perceptual attributes (such as pitch, temporal grouping, timbre, spatial location etc.) which are processed faster and with substantial or complete commonality between different individuals, and proceeding through successively more complex and abstract levels of processing, which become more idiosyncratic to the specific training and experience of an individual listener. Thus the ‘cultural meaning’ of a piece of music, which is regarded as being at the remotest level of this sequence, is conceived of as being the most personal, idiosyncratic and
unpredictable (or even inexplicable) aspect of the whole experience, while the basic perceptual attributes are a more tractable proposition for empirical and theoretical accounts. A variety of kinds of evidence are offered in support of this view: speed of processing (a simple decision task on a basic perceptual attribute will be made faster than on a more ‘complex’ attribute), developmental sequence (children seem to acquire a sensitivity to simple perceptual attributes before more complex attributes), and inter-subject stability (judgements of basic features such as the grouping of events typically show much more agreement between subjects than do judgements of music’s ‘semantic content’, for example). The empirical truth of these findings, however, does not necessarily entail the interpretation provided by cognitive theory, and furthermore in an enculturated adult developmental sequence or relatively small differences in speed of processing may be of little or no consequence for perceptual experience. The view taken in this paper is that whatever the ‘flow of information’, in effect more ‘abstract’ levels are just as directly available (and paradoxically possibly even more so) as are those regarded as basic and primary.

Ecological perceptual theory

An alternative view of perception and cognition has existed for the best part of 50 years, has been almost completely ignored by the psychology of music, but offers a radically different and extremely fruitful outlook. This is the perceptual theory of J. J. Gibson, often referred to as an ecological or direct realist approach. For the purposes of this paper, I will adopt just three fundamental principles from ecological theory: i) stimulus information is highly structured and specifies its source directly; ii) source specification is an aspect of meaning; iii) while ecological theory may have been developed primarily in relation to the perception of the natural environment, there is no sharp discontinuity between nature and culture, and the manner in which cultural meanings are available to a perceiver is no different from the way in which natural meanings are. Gibson himself pointed out that culture is as dependent on material reality as is the natural environment, and that our response to the cultural environment is similarly dependent on the ability to pick up perceptual information: “Symbols are taken to be profoundly different from things. But let us be clear about this. There have to be modes of stimulation, or ways of conveying information, for any individual to perceive anything, however abstract. He must be sensitive to stimuli no matter how universal or fine-spun the thing he apprehends. No symbol exists except as it is realized in sound, projected light, mechanical contact, or the like. All knowledge rests on sensitivity.” (Gibson, 1966, p. 26).

Gibson developed these ideas primarily in relation to vision, and there has been only a rather slow attempt to apply the principles to hearing, and music in particular (though see Balzano, 1986; Bregman, 1990; Dowling & Harwood, 1986; Handel, 1989; Windsor, 1995). A great deal stands to be gained by doing so, however, and particularly in bringing some kind of continuity and unity to the way in which we understand the perception of ‘basic’ features and cultural meanings of sounds. The ecological approach asserts that the source of a sound is directly specified in the stimulus information as it arrives at the sensory system of a perceiver: the spatial location, the material (specified in the frequency relations that we identify as timbre), the mode of excitation (specified in the attack characteristics), its separation from other sources (specified in the dependence or independence of behaviour of frequency components, vibrato, dynamic variation, etc.), and so on. Bregman (1990) gives a thorough and comprehensive account of this aspect of auditory and musical perception from a broadly ecological perspective. But the principle can be, and has been, extended into far more apparently rarified areas. Windsor (1995) has pointed out that just as sounds specify directly the invariants of the natural environment, so too do they specify the constancies (in fact we can call them ‘invariants’) of the cultural environment. The resistance to taking this step has been twofold: i) higher level properties of music (such as tonality and metre, for example) are held to be mental constructs rather properties of the material itself; ii) cultural meaning has, with the influence of linguistics and semiotics, been theorised as based on arbitrary codes,
and thus subject to constant and arbitrary change. Both assumptions are unnecessary. The tonality or metre of a piece of music is as much a property of that music (in relation to a perceiver) as is the identity of the particular physical instrument(s) playing it - and is as shared with all the other music that shares the tonality or metre as is the instrumental commonality of all the other music played by the same instrument(s). Tonality can be regarded as directly specified in sounds (though obviously distributed over time) in just the same way that the roughness of two pieces of sandpaper rubbed together is specified in the acoustical information. Similarly, although it is true that the coding of a cultural unit with its conventional meaning is arbitrary (in the sense of not being determined by the form or substance of the signifier) and thus potentially infinitely changeable, the reality is that once that cultural unit is embedded within a system of any complexity, its meaning is effectively as invariant as a natural law. It is quite possible in theory to decide that the signifier <<tuba>> will from now on mean (i.e. denote the concept) ‘violin’, but essentially impossible in practice to effect such a change, given the weight of human culture. Thus <<tuba>> means ‘tuba’ for a particular cultural community with about the same degree of flexibility that the acoustical information broadcast by the slamming of a car door specifies just that - a car door slamming.

An example
From this theoretical base I let me present a brief but suggestive musical example which illustrates how the view proposed here might be used. It is a rock guitar solo - specifically the rendering of The Star Spangled Banner given by Jimi Hendrix at Woodstock in 1969. In the absence of a sound example to accompany this version of the paper, let me provide a sketch description of the recorded performance (which can be heard on Polydor CD 517 235-2 Jimi Hendrix: The Ultimate Experience). It is a live performance, essentially a guitar solo for Hendrix, though with occasional accompanying interjections from the drummer and bass player. The melody of the anthem itself (there is no singing) is introduced by high guitar feedback and then played relatively straight as far as the pitches are concerned for the first four phrases (to 0:32”) and their repeat (to 1:06”), although with increasing amounts of pitch bending and embellishment. The sound in both of these first two large phrases is characterised, however, by constant incipient feedback that is only just kept in abeyance, a high level of distortion, and the swirling spatial effects of using Lesley loudspeakers (loudspeakers with physically rotating cones). The second section of the theme (from 1:10”) is treated to a disintegration into feedback and roaring swoops of sound up and down the guitar. The rest of the performance (which lasts for a total of about four minutes) consists of a mixture of dense distorted guitar sound and feedback, with sudden islands of clarity where the outlines of the melody can be heard continuing - as if the anthem is intermittently submerged under the sounds of the guitar, to be glimpsed from time to time. Finally, as the melody is swallowed up once again by this mass of sound, it is transformed into the opening of “Purple Haze” - which might almost be regarded as Hendrix’s signature tune.

Three kinds of instability, ranging from cultural practices and the individual identity of the performer to sound itself, are specified in the sounds of this performance, as Figure 1 attempts to summarise, and do not differ in their immediacy - even if they require different durations of material to become audible. It may be quicker to identify the opening sounds as specifying the guitar than as specifying Hendrix, or the Star Spangled Banner, but this is not because the individual identity of Hendrix, or The Star Spangled Banner are more abstract invariants than the guitar, but because more perceptual information is needed to specify them. There is no distinction in the immediacy, or directness, with which either is specified, however, for an appropriately attuned listener.
I. Factors which specify unstable/antagonistic cultural practices:
   i) Pitch pattern => American National Anthem => Nationalism
   ii) Instrumentation => Rock, Youth Culture => Resistance
   iii) Timbre, characteristic devices => Jimi Hendrix => Black America

II. Factors which specify instability in the musical material
   i) Pitch Bend => The breakdown of the discrete/continuous distinction in pitch.
   ii) Disintegration into 'pure material' (arpeggios) => breakdown of melodic identity
   iii) National Anthem/Rock Gestures (e.g. trilling thirds)

III. Factors which specify instability in the sound
   i) Distortion (electrical - amplification gain) => Instability of pitch & timbre
   ii) Sound rotation (Lesley speakers) => Instability of location
   iii) Feedback (gain) => Instability of instrument and mode of production.

As the figure attempts to show in very summary form, the meaning and critical value of this performance (and I acknowledge that the same perceptual information can specify different meanings to different audiences - a quite explicit aspect of mutuality that is inherent in Gibson's ecological view) are directly specified at a number of levels simultaneously, and all with the same directness and immediacy. Culture and ideology - which are those aspects of meaning of which listeners are primarily aware - are just as material (in the concreteness of the practices in which they are embodied) as are the instrument that is used to generate this performance, and as perceptual sources they are just as much a part of the environment.

Summary. Prospects. Critique

This paper has proposed that while aesthetics and critical theory on the one band, and the psychology of music on the other, may have somewhat aims and agendas, there is considerable scope for a far more fruitful dialogue between them than has been witnessed for most of the period of the 'modern' psychology of music. The ecological approach to perception, and its espousal of a form of direct realism, offers an important way to develop such a dialogue by proposing a theory which draws together the perception of supposedly 'basic' properties and supposedly 'abstract' characteristics. By emphasising the fundamental importance of perceptual information, and the manner in which perceptual information specifies sources (ranging from physical sources to cultural systems) the perceptual immediacy and materiality of what have previously been regarded as completely divergent qualities of music are underlined. It is remarkable that while the psychology of perception and critical theory have commonly been regarded as being poles apart, the direct realist approach brings them into very close proximity. When Adorno writes that "What an artist has to say is said through figuration; it is never a message carried by figuration" (Adorno, 1984: 216), he expresses the materiality of music's critical content, just as Gibson (see above) emphasises the concreteness of both nature and culture.

The brief analysis presented here might justly be criticised for dealing with the music concerned in a manner that is no less simplistic than that offered by more standard cognitive accounts. It is impossible within the confines of a paper such as this to demonstrate anything more than the broad outlines of the approach. Two indicators, however, point to the potential of such an approach: first, more detailed analyses based on the same or similar principles exist (e.g. Windsor, 1995) which show something of the scope of the approach; second, even in the cursory analyses contained within this paper, a much broader and more diverse range of attributes is tackled than is usual, encompassing issues that have conventionally been regarded as beyond the bounds of music perception.
The issue is not just one of theory (psychological, musical, or cultural) and analysis, however. As far as empirical investigation is concerned, a crucial question becomes "What are the invariants that specify...?" - and this can be 'aimed' at a whole variety of levels - from motifs to ideologies. Note also that this does not invalidate or go against a great deal of existing work in music perception - particularly the kind of research in auditory scene analysis (Bregman, 1990) which is explicitly influenced by ecological acoustics. It does, however, radically change the emphasis from internal processes to external realities. Finally, although contemporary connectionism is an enormously over-simplified view of either the mind or the brain, it nonetheless is consistent with the ecological approach adopted here, and demonstrates in a concrete manner how a system that does not make use of representations, but is shaped by exposure to a structured environment, can behave in a complex and interesting manner.

References
The Perception of Associative Structure in Atonal Music

N. Dibben,
Music Department, University of Newcastle, UK

E. F. Clarke,
Music Department, University of Sheffield, UK

Introduction

It has been widely proposed, or assumed, within both music theory and the psychology of music that listeners hear relationships between abstract underlying structures in music, as well as surface relationships, and that a hierarchy of tonal structures is fundamental to the listening experience. In this respect, as well as in others, a parallel with modern linguistic theory may be drawn, with underlying structure playing a role that is not too dissimilar to the 'gist' of an utterance (Sloboda, 1985). Theory of this kind has developed primarily in relation to tonal music, with the work of Schenker (e.g. Schenker, 1979) and more recently Lerdahl and Jackendoff (Lerdahl & Jackendoff, 1983) being the most obvious examples. More recently, a similar approach has been proposed for the post-tonal music of this century (Lerdahl, 1989), using related (though adapted) reductional principles to derive underlying pitch structures in the absence of an organising tonal system.

Empirical evidence for or against reductional theory is surprisingly thin on the ground (e.g. Bigand, 1990; Dibben, 1994; Serafine, Glassman & Overbeeke, 1989), but from this limited literature there is more evidence for listeners' ability to identify reductional structures in tonal music than in atonal music. There may be a number of reasons for this: i) listeners tend to be much more unfamiliar with atonal music and may have difficulty in making stable perceptual judgements about any aspect of its pitch structure; ii) the specific reductional theory in Lerdahl (1989) may be flawed, and listeners may be aware of reductional structures in atonal music, but not of the type proposed by Lerdahl and investigated by Dibben; iii) the experimental methods used by Dibben (1994 and 1996) may be at fault - due to problems of memory and attention; iv) despite the intuitions of those theorists who have proposed extensions of reductional analysis from the tonal repertoire into atonal music, it may simply be the case that reductional structures do not play a role, or play a drastically attenuated role, in listeners' responses to atonal music. Whilst acknowledging the possibility that one of the first three explanations offered above may be correct, it is this final possibility that we wish to examine here.

The alternative perspective that we propose is that associative structure plays a far more significant role than has previously been recognised, and in a way that is rather different from that offered before. Although there is little empirical evidence for listeners showing any sensitivity to reductional structures in atonal music, there is far too much informal evidence that listeners hear some kind of structure and unity in this music simply to dismiss their response as entirely unstructured. Qualitative data from interviews with performers (Dibben, 1996) suggests that the sense of connection and unity that they experience in atonal pieces with which they are familiar is far more attributable to associative connections between elements of the music than to abstracted underlying structures - which they do not spontaneously mention at all.

Associative structure in music has lived in the shade of hierarchical structure - both as far a theory is concerned and in terms of empirical research. Meyer (1973) provided a stimulating discussion of a variety of issues in the associative structure of melody, but there has been little attempt either to follow this up, or to investigate such issues through experiments (e.g. Meyer & Rosner, 1981; Pollard-Gott, 1983). The approach to associative structure proposed here differs from that explicitly presented in Meyer in that it encompasses associative links in two axes, which are somewhat similar to the syntagmatic and paradigmatic axes of Saussurean linguistics. Meyer primarily
discusses the syntagmatic axis, analysing and categorising the associative links between elements occurring within the same piece of music but displaced in time. The paradigmatic axis (Saussure himself originally called this the associative axis) is concerned with the relationship between any instance of an element and the other elements belonging to the same category with which it could be substituted. What this axis is intended to capture in the theory as we present it here is the manner in which musical materials refer beyond themselves to archetypes, prototypes, or simply 'other instances' that may not be present here and now but which give a basis and context for evaluating those that are. The principle arises as an extension of the idea that the atonal repertoire - particularly that of the free atonal music of the Second Viennese School - still contains many characteristics (of phrasing, motivic structure, rhythm, texture, etc) that are powerful legacies of the tonal period, and that these associative or gestural links seem to play a powerful role in listening. In sum, we propose that listeners make associative links between musical elements that are actually present in any given piece, and at the same time make associations with similar or functionally equivalent elements or gestures in a far wider repertoire of music with which they may be familiar. These associative links beyond a particular piece play an important referential role, and also mediate associations within any given piece that may not seem obvious or even likely. For example, a gesture in an atonal work that is associatively linked to the category 'mordent' may become associated with a modified turn figure later in the same piece by virtue of their common membership of the gestural class "ornament" - even if on the surface of the piece in question they may bear little apparent relationship to one another. More significantly, perhaps, elements may acquire a functional significance by virtue of associations of this kind that goes far beyond their apparent properties: an atonal chord configuration that has no harmonic qualities that could give it a cadential role may acquire a cadential function by mimicking some superficial aspect (rhythmic, textural, voice-leading) of a tonal cadence and thereby becoming associated with its closural function. This is, in essence, an argument for paying more attention to the historical nature of musical material - rather than simply treating it as raw materials in the here and now. Taking this a step further, these same associative links also lead beyond what is commonly regarded as the domain of music into a more general system of cultural reference, as we illustrate with a brief analytical example at the end of this paper.

Methodology

An immediate question is whether it is possible to investigate this theory empirically, and if so how. The idea that musical elements associate with functional categories suggests that some kind of priming methodology might be appropriate. The hypothesis would be that if a motif associates with a general functional category of some kind, it should prime this category, and thus prime the pick-up of any subsequent instance of the category (provided that the category is still active), the extent of this priming being related to the prototypicality of both first and second instances. However, in the light of a variety of methodological problems in trying to put this empirical method into practice, we have turned to a rather more descriptive and qualitative approach as a more feasible initial exploratory study.

Experiment

The experiment presented here is an attempt to assess the degree to which associative links play a role in listeners' responses to a piece of atonal music with which they are reasonably familiar, and the nature of such associative links. It consisted of a familiarisation phase plus the experiment proper, the music used being the fourth of Webern's 5 pieces for String Quartet. In the familiarisation phase, each subject was given a tape recording of the piece and was asked to listen to it approximately twice a day for 7 days. Subjects were told that at the end of the week they would perform a short listening task but were not told what this would entail. Eight undergraduate music students participated in the experiment, and provided information about their familiarity with atonal music, whether they had heard the piece before, and the number of times they had listened to the tape.
The experiment itself consisted of two sections: a listening task and a discussion session. For the listening task subjects were instructed to listen to the piece and to tap on the desk in front of them every time they heard something in the music which formed an association or a connection with anything else, and to indicate anything which gave the piece coherence. Each time the subject tapped, the interviewer marked the corresponding place on the score. The piece was played twice in succession with the same procedure each time in order to allow the subject to add to or amend taps made on the first hearing.

Following these two complete playings, the marked score was placed in front of the subject and the interviewer played the recording once more, this time pausing at each of the marked positions. The subject was then asked to indicate exactly where they had meant to tap, what the tap referred to, and to mark any associations or connections by circling the appropriate parts on the score and linking them with lines to show the connectedness of different musical ideas. The discussion session was recorded onto tape.

Results

Previous familiarity.

While all subjects were familiar with the atonal repertoire as stated above, only 2 of the 8 subjects said that they listened to atonal music frequently. Only one subject had heard the movement before, but was relatively unfamiliar with it having heard it only twice before a long time previously. The average number of taps made by subjects, and therefore the average number of connections identified, was 7.31. The only effect of familiarity was a positive correlation between listeners' familiarity with the piece and the number of connections identified on the first hearing (Spearman Rank Correlation, p=.05): increased familiarity with the piece allows listeners to identify more connections.

Types of connection

A total of (approximately) 31 different connections were identified, encompassing virtually every note of the movement. Certain parameters were mentioned more than others in the discussion of heard connections: timbre (23), texture (18), figure/pattern (17) and register (17) were cited most frequently. Less common were mentions of melodic contour (9) and exact pitch repetition (8). Although pitch repetition was mentioned as a factor a small number of times, it was the least mentioned factor. It is significant that the only occasions on which pitch repetition was mentioned were extremely local-level connections, i.e. events in close proximity at the musical surface - such as the repeated Eb cello note in bars 1-2 and the viola E-F# dyad in bars 2-3. This is consistent with Deutsch's (1977) demonstration of the stability of short term memory for single pitches. Nearly all listeners remarked explicitly on the lack of importance of exact pitch repetition to their conception of the piece. While exact pitch repetition may be unimportant however, pitch height (register) appeared to play a far more significant role. Not only was it mentioned more frequently, but it also formed the basis for larger-scale connections, such as that between bars 1-2 and 11-12.

Conclusions

The findings of this exploratory study suggest that while the pitch content of events at the musical surface may be significant in the formation of local level connections, at higher levels, i.e. distant associations across the musical surface, pitch is far less important than is register, timbre and texture. Indeed, the subjects in this study never spontaneously mentioned pitch regions of any kind. Thus, these results would seem to suggest that listeners' representations of atonal music are not captured by a reduced pitch representation since such a representation excludes the direct representation of factors such as texture and timbre.
A further aspect of the subjects’ discussion of this music is their allusion to ‘functional gestures’. At a number points, subjects mention that a figure or element seems to play a particular role in the music, and sometimes is even connected to another figure/element playing the same role but which may have quite different properties. For instance a number of subjects remark on the closing function of the pizzicato chords in bars 2 and 12, and one subject also mentions the similar function she attributes to the final separated pizzicato note in the viola in bar 9 and a sense that something around the end of bar 5 seems to perform the same function (possibly the final isolated cello note at the beginning of bar 6). A second example is the remark made by two of the subjects that figures in the music seem to have something like an ‘antecedent-consequent’ relationship. In one case (bars 1-3) this involves very different musical materials, while in the other (bars 4 & 6) it can be attributed to an inversion of the melodic contour. A third instance, and one which is additionally interesting in its reference to extra-musical issues, is the reference by two subjects to the clock-like character of the viola ostinato in bars 7-9 which subjects describe as ‘winding down’ as bar 10 is approached.

Finally, there is a striking reference to both the musical and extra-musical qualities of the lyrical violin melody in bars 7-9: the subject describes this as having a strong character of ‘longing’ about it. It is striking that the melody consists of repeated descending minor third figures, which together with the metrical orientation and hairpin dynamics give the melody a character which is strongly reminiscent of a more conventional tonal ‘sigh’ motif. We see this as an example of the kind of historical reference discussed above.

While listeners’ remarks on the reasons for connections gives some indication of the causes for the formation of connections in this movement, the types and frequency of causes will differ according to the material of each individual piece. The reasons for the connections may also be influenced by subjects’ particular musical training and the extent to which listeners have access to what may be procedural rather than declarative knowledge is debatable. For these reasons similar research needs to be undertaken with other atonal pieces to determine the extent to which these findings are generalisable.

The example below is an illustration of the way in which the associative ideas outlined above might apply to a short atonal piece (Schoenberg’s piano piece op.19, no.6). There are a number of similarities between this analysis and reductational analyses. For example, the brackets below the score indicate a hierarchical segmentation based on the grouping principles as outlined by Lerdahl and Jackendoff (1983). Similarly, the tree diagrams above the score represent hierarchical relationships which differ from reductional models by being isolated structures restricted to local levels and to contexts in which they have an association with their tonal counterparts. Taken together, these constitute an example of one of the functional gestures discussed above. The circles identify associations between tri-chords whereas the boxes identify associations between the more linear elements of the music. These undergo a transformation from a neighbour-note figure, to an incomplete neighbour-note, to an elaborated neighbour-note, and finally to an incomplete neighbour-note displaced by an octave. Whereas the association between the circled material (tri-chords) is an entirely intra-opus connection, those between boxed material (linear motifs) are both intra-opus and extra-opus by virtue of their reference to tonal music. In addition to these musical associations there are also cultural references of a more general kind which are focused by the context in which this piece was written: Schoenberg wrote it as a hommage to Mahler on his death. The neighbour-note motive, and its transformation into something more reminiscent of an appoggiatura (particularly in b.7), carries with it an association with sighing, and the particular sonority of the tri-chords with their particular spacing and register carries an association with bells tolling.
References
Interpretative Criterion and Understanding of the Textural Hierarchy
A Production-study about some internal representations
Favio Shifres - Universidad Nacional de La Plata - Argentina

Introduction
When the performer translates a score into a sound musical experience, he applies an action system which is his Interpretative Criterion (IC). This set often shows deviations from the written score. Many studies (e.g. Sundberg, Friberg & Frydén, 1991; Shaffer & Todd, 1987, Gabrielsson, 1987) reveal that these deviations are expressive, and usually rule-governed. Thus, the expression can be related to the musical structure (Clarke, 1985; Shaffer, 1995). Furthermore the rule-based system would have some correlates in the perceptual input and this would allow the listener to interpret the deviations and to rebuild the performer’s expressive intention (Palmer, 1989), founding codes to build the communication of the musical idea between both of them (Clarke, 1985). Through this double code system, the musical structure is revealed to the listener.

The message can communicate both structural features -as meter (Sloboda, 1983), phrasing (Fryberg, Sundberg & Frydén, 1987), rhythm's regularity (Clarke & Baker-Short, 1987)- and emotional characters (Gabrielsson & Lindström, 1995; Gabrielsson & Juslin, 1996). Thus, the performer mediates in the listener’s internal construction of very complex factor. May be the more complex the attribute to communicate is, the more important the role of performance will be the role played by the performance will be.

The work itself, does not reveal to the listener all of its own expressive content without the performer’s mediation. The high level of responsibility assumed from this forces the performer to insure the IC Relevance and Efficacy. The systematic analysis of the piece (Berry, 1989) and the study of the time and author stylistic characteristics allow the performer to notice the logic of the musical content in order to (there seems to be a general agreement on this point) reveal to the listener as clear as possible. A IC is relevant when its logic fits into the logic of the structure and it is effective if it can communicate the performer’s expressive intentions. However, when we think about more sophisticated aesthetic choices, Relevance and Efficacy are dialectically related. As a result, an effective IC can be irrelevant in that sense because it is either exaggerated, obvious, or banal.

The texture is a high complexity organisational factor. The interpreter finds hierarchies among the components of the structural organisation through the analysis of the work (e.g. the distinctness by peculiar qualities of a figure-background relationship) in order to guarantee the IC relevance. The performer can operate on the perceptual hierarchy of these lines by a number of actions, which being ruled-governed, would be unconscious, as the chordal asynchrony (Palmer, 1989) - insuring the IC efficacy. Nevertheless an exaggerated performance, even effective, is not relevant.

The skill to focus a melodic line of the texture is related to the Gestalt principles (Sloboda, 1985). When complex textures are analysed, this principles can apply not only to the focal line configuration but also to all the components concurrence with their particular dependence-interdependence relationships (Berry, 1987). Thus, this skill would be governed by the features of the musical structure. However, real life music is much more complex: the textural structures are often either configurations that are far from gestalt models or different
conflicting concurrent dispositions, since the interdependence weaving relationships among the
textural components is usually very complex. Consequently, the textural organisation itself
often prevents the listeners from representing the hierarchy. In those cases, the IC would be
crucial (Cook, 1990).

Furthermore, in the analysis of the representation of the textural organisation, two different
situations should be distinguished: the first one takes place as a result of the active search of
the given configuration; and the second one occurs as rising from the passive awareness of
such configuration (Sloboda, 1985). That active search could be caused both by the previous
knowledge, -not only of general musical background, but also of the particular piece-, and by
some kind of task prior to the listening, which could catalyse said task.

This study is a part of a research which explores the nature of the internal representation of the
texture related to the IC by studying the listeners' skill to isolate the figure from the
background while listening to complex textures, in first listening contexts.

**Previous Research**

Four excerpts from nineteenth century piano repertoire were chosen according to a number of
categories for the textural analysis. They were recorded by a pianist (ADR) in two versions: 1)
making the melody perceptually salient according to his own analysis (version A), and 2)
deadpan performance (version B). Records were analysed by 5 experts - professors of the
Faculty of Fine Arts, professional musicians who have acknowledged background as either
choir conductor, orchestra conductor or music educators, all of them are pianists and have
acknowledged expertise in aural skills and assessment of instrumental performances -. They
chose the best version for the experiment. The agreement among them and the pianist intention
was: 83% for the example 1, and 100% for the example 2.

In the first experiment (Shifres, in press) listeners had to answer: *Which is the version where
the melody (figure) is clearer?*. For the ex. 1, the 29% chose version A, 24% version B, and
47% could not make a difference between them. For the ex. 2: 68% chose version A, 13%
version B, and 19% could not make a difference. The second study (Shifres, 1996) examined
the efficacy of this differentiation, considering a performance as effective when it enables
listener to build the expected hierarchical internal representation. Listeners had to answer:
*Could you identify this melody (figure) while you were listening to?*. Answers were classified
in 10 ordered categories showed medians of 10 and 9 (NS)to versions A and B respectively of
the ex. 1 and 9 and 3 ($p < .001$).

**Method**

The present study relates more direct information entailed to the internal representation of the
textural hierarchy. For this a production test was made using an adaptation of the *Aloud
Thinking* technique. Here, subjects give a “quick description” (as a response in musical terms)
of the process that they are using as they are doing the experimental activity. It is assumed that
a careful exam of these responses allows us to obtain some access to the subject’ thinking
pattern, since “the more interesting features of musical understanding may be expressible only
in a musical medium” (Clarke, 1985. p.209). Records of a version (A or B) of each example
were presented twice, separated by the same time span as the example. Subjects had to listen
to the first one focusing on the main melody (figure) and had to sing (or whistle) it, over the
record, while they were listening to the second one. They were asked to show this voice
(“thinking aloud”) as they were listening to it, and to give as much information as possible.
Responses were recorded in order to their subsequent analysis. Each test took 3 minutes. No subjects declared to know the heard excerpts.

Selected musical stimuli were the same as the previous tests (see Previous Research). Subjects were sorted in 4 groups in order to reduce the amount of listenings and to shorten the test length - in order to avoid tiredness and have a positive attitude to the task throughout the test. Each group listened to the version A of two examples and the versions B of the other. Common version were taken as a pre-test to measure the reliability for the responses to the other versions. (Responses given to common versions showed in every case no significant differences). 195 applicants for admission at the Faculty of Fine Arts of the UNLP (1996), sorted in two categories according to their previous musical experience were taken as subjects for this test.

Results
In order to avoid a wrong interpretation of data, only responses that can show some property of the study problem must be taken into consideration: if the subject does not do any performance or if he carries out an "undesciphorable" performance, one can not consider that response as an internal representation indicator, but simply as inability to solve the task. However, it is possible to take some information from a "real" performance of the subject. Thereupon, not only the responses quality -in terms of metric-rhythmic configuration and melodic-tonal organisation- were taken into account, but also other attributes of the performance -such as articulation, phrasing, dynamics, all of them revealing coherence and fullness of meaningful musical units - were analysed.

An open observational system of data record and storage was used: every response was listened to and a verbal description of each was done. Those descriptions were classified in a seven category system. Responses were analysed by trained people. In order to reach the maximum agreement, inter-ratters discussion sessions were held when necessary. The system was improved by rewriting some of the categories according to those analysis. Then, responses were recategorized according to this adjustment.

The isolated categories were: (The response shows....)
1. The main melodic line (figure).
2. Some excerpts of the main melodic line (figure), which form meaningful units.
3. The main melodic line (figure) - either complete or incomplete - with some added sounds belonging to other textural components (background).
4. Another textural line (background) is noticed in the response.
5. The complete basic (minimum) rhythmic beat (background) with pitches of the main melodic line (figure).
6. No identifiable (clear) line (unconnected pitches, voice leading non-existent, non explicit contrame洛d, a line tonal and/or rhythmically non configured, etc.).
7. Nothing (the subject loses immediately the continuity, he tries but does not emit any sound, he only emits very isolated sounds at the excerpt boundaries, etc.).

Findings from two of the test examples are reported here:
Example 1: Schumann's *Noveletta No. 6*

![Example 1](image)

Example 2: Schumann's *In der Nacht op. 12 No. 5*

![Example 2](image)

No response of these examples was classified into the fourth category. Therefore, observed frequencies (in percentage) of the other categories are introduced, in order to compare data arising from both versions. (Table 1)

<table>
<thead>
<tr>
<th>Example</th>
<th>Category 1</th>
<th>Category 2</th>
<th>Category 3</th>
<th>Category 4</th>
<th>Category 5</th>
<th>Category 6</th>
<th>Category 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version A</td>
<td>41</td>
<td>8</td>
<td>33</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Version B</td>
<td>44</td>
<td>15</td>
<td>23</td>
<td>5</td>
<td>4</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>13</td>
<td>27</td>
<td>11</td>
<td>0</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>5</td>
<td>41</td>
<td>11</td>
<td>26</td>
<td>11</td>
<td>11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example</th>
<th>Analysable</th>
<th>Non Analysable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version A</td>
<td>86</td>
<td>13</td>
</tr>
<tr>
<td>Version B</td>
<td>87</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>93</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>63</td>
<td>37</td>
</tr>
</tbody>
</table>

| $\chi^2$ | 5.32       | 0.00           |
|          | NS         | $p < .001$     |

No differences among subjects with a different level of previous musical experience were found.

**Discussion**

The used procedure enabled to reduce memory, tests design and chance incidence compared with other tests. Moreover, subjects’ commitment was higher (this topic is crucial in our work environment). Each response can be analysed in musical terms in order to deepen the study of the representation nature, which is an important advantage from the other. However, it shows some disadvantages: 1) Musical expertise degree involved in the response, knowing that

Table 2
sample of the internal representation is intervened by the task difficulty itself when it engages a performance activity; 2) Difficulty for objective assessment of the responses; 3) Limit to search non analysable responses. Nevertheless, findings for example 2 could be talked about IC’ incidence in the subject’ song response organisation. As these results are highly coherent with the previous tests’ findings, which did not have these objections, it is useful to interpret all of the data together.

The percentage of responses classified into the category 1, would reveal that, in the example 1, the main melody is focus regardless of the performer’s intention, while the performer’s mediation rises as decisive to the example 2. This appears to be against intuition, since a non sophisticate description would predict the opposite: the melody is the highest level, being the other features very similar. Melody description, in terms of the Gestalt principles, can not give a reason for this difference by itself either. However, there are some characteristics that can make great differences between both of examples: 1) the tempo would contribute to configure the melody (Shaffer, 1995); 2) Rhythmic relationship between figure and background (1:4 /\ - and 1:2 -\ /\ in example; and 1:8 -\ /\ - and 1:4 -\ /\ in the ex. 2); 3) accompaniment directionality that shows gestures forward the melody, in the second example. 4) The bass line, in the second example is reinforcing the upbeat of each time. Thus, the conflict is noticeable in the many textural factors concurrence. These can interact either reinforcing or neutralising their reciprocal effects and, therefore demanding a more o less emphatic performance.

High responses percentage for the example 2B category 3 could be understood by the same explanation. Most responses classified in this category just added that background upbeat to the figure. Consequently the necessity of deepening the study by a more detailed description in musical terms of each category surfaces. Besides, this fact gives notice about the specific character of the used classificatory system for these examples.

It is possible to think that the IC for both version were not different, being the version B non deadpan. The IC was evaluated in three levels: the pianist, who was evaluated by himself in the same way as he did for the example 2; the panel, who did not reach a complete agreement while evaluating the two versions even though it could find differences between both of them; and the listeners. “(T)he temporal nature of performance creates priorities different from those of the analyst, for whom time becomes converted into space - or at least loses its dynamics and irreversible quality” (Clarke; 1995 p. 25). In this way, the performer approaches the analyst’s view, since the multiple repetitions of the excerpt during the study process and the non perceptual knowledge of the configuration decreases these qualities when he is making the internal representation of his own performance. May be, this is why his representation is so different form the passive listener’s. According to this, the real temporal nature of music would be only experienced by the listener.

Why does the difference among the three levels take place with only one example? Probably, this difference could be explained by the description of the textural features again. These characteristics would influence the way to organise the structural properties (Clarke, 1985. p. 210) and to carry it out through his IC.

We face again with the fact that the traditional analysis models of the texture do not explain the phenomena involving the structure. A satisfactory model must take into account the temporal nature of this communicational phenomenon, beyond simple local properties of the structure.
To ensure the IC relevance and efficacy demands an interpretation engaged not only with the analysis of the piece but also with the listener’s representation. Therefore a descriptive theory of texture with the view to drawing interpretative actions must take into account of: 1) both its synchronic nature and its diachronic development; 2) dynamic and irreversible quality of the passive listening; 3) configurations of the textural components in the Gestalt terms; 4) the dependence/interdependence relationships among components; and 5) the incidence of the tonal and metric elements.

References
The question of universal (natural) versus conventional or innate versus learned in music formation and perception, has, as we know, engaged many researchers in exploring various phenomena. Here we explore the use of these concepts regarding principles of organization — which are considered schemata — and their meaning to the listener.

How arbitrary are the schemata in music? We shall try to determine this in order to understand more thoroughly the rules of musical organization and how they connect with a specific cultural stylistic ideal. We shall further try to find out why should certain schemata be chosen, consciously or not, in certain cultures or eras.  

Meaning and the Stylistic Ideal

We assume that musical styles vary throughout cultures and eras due to different stylistic ideals and not because the meaning of musical rules is susceptible to change. It seems that musical meaning relates to mental experiences that may be manifested in several domains, each them confined between two extremes. The main (interrelated) variables by which we examine a stylistic ideal are the following:

* Arousal: Generally speaking, two poles can be distinguished: calm and excited, the latter containing an abundance of possibilities which gained attention (e.g., Sloboda 1991; Gabrielsson and Juslin 1996).

* Connection or disconnection to the outside world, by which we mean music composed with predetermined connections to certain non-musical factors, such as functional music (common outside the West) or "programmatic" music. This factor is responsible for dispute among musical approaches (e.g., "referential" versus "absolutist"). However, we shall regard it as reflecting the stylistic ideal.

* Nature of flow (static or dynamic) can be defined in terms of directionality (momentary or overall; clear or suspended) and complexity (momentary or overall) (Cohen 1994).

Different ideals set preferences as to the meaning or emotional state suggested to listeners. How do schemata affect these preferences?

Schemata

The concept of schemata has drawn the attention of many researchers from domains outside music and has been given many names ("prototype," "archetype," "conceptual model," etc.). Bartlett (1932) was the first to talk about dynamic schemata in psychology. Today the concept is also used in cognitive musicology. We consider a schema to be an abstraction or principle that has been conceptualizing in our minds, consciously or not, and which is used for the mental representation of information. Most of the schemata are principles of musical organization on hierarchical levels and consider "raw material" (e.g., scales) and compositional rules. Then a certain schema can be regarded as a realization of another schema on a more abstract level. Thus the schema, which is the main factor in producing directionality, connects various musical events and arouses expectations, both of which are essential for the creation of meaning.

1Unfortunately, due to a shortage of space, we are unable to provide musical figures and a reference list in this paper. We would be happy to supply both upon request.
Learned and Innate Schemata

It is customary to distinguish between rules of organization that are learned (scales, chords, forms) and vary throughout cultures and eras and rules that are "automatic," so to speak, natural innate principles that can be observed in contours (Dowling 1978; Edworthy 1985), theories of Gestalt, grouping rules (Lerdahl and Jackendoff 1983), rules of expectations (Meyer 1973; Schmuekler 1989), and so forth (for a summary of the effect of innate schemata see Boltz and Jones 1986). The confrontation between learned and innate, or natural, is conspicuous in pairs of contrasting terms: "bottom up/top down" (Narmour 1991); "musical alphabet"/Gestalt rules (Deutsch and Feroe 1981); intervals/contours (Dowling 1978); etc. However, we feel compelled to connect the above distinction to specific schemata — learned or natural — chosen because of the stylistic ideal.

Learned schemata

These refer to man-made organizations (such as scales, chords, harmonic patterns, rhythmic patterns, specific timbres, or forms). They are not known outside the realm of music. Learned schemata vary between cultures and eras; they are chosen cognitively, and most of them receive quantification (intervals, rhythmic patterns, etc.). Most researchers still assume learned schemata are the result of conventions, thus ignoring significant studies (e.g., Balzano 1980; Agmon 1989; Brown 1981) that discuss choice of learned schemata related to pitch parameters in Western tonal music: the choice of 12 chromatic tones; of the diatonic system of 7; major and minor scales; the interval system; and the chords. These studies show that the above schemata are far from being arbitrary: they fulfill demands of utmost efficiency and coherence. Moreover, these studies reveal the existence of other hypothetical systems that fulfill the same criteria but fail to meet psychoacoustic constraints. Moreover, the same coherence demands underlie the desire of Western ideal for a complex organization on a long run, so that a musical piece may be looked at as a well-cut diamond, separated from its surroundings. However, in non-Western cultures, which clearly obey different ideals, scale-interval systems fail to achieve coherence (Cohen 1994) or categorization (Keefe et al. 1991) and are not fit for clear overall organizations, but tend to focus on whatever is momentary.

Furthermore, while the Western "solution" in selecting raw material seems to be a single one, other solutions may be realized opulently, within certain limits. Therefore, the term "alphabet" (Deutsch and Feroe 1981), which designates a somewhat arbitrary choice (contrary to the rules of language) will not do for describing learned schemata.

Classification principles of learned schemata

The main criterion for classification regards the contribution of schemata to the various types of complexity and directionality (Cohen 1994). The factors that seem relevant to us for establishing this (apart from the demand for coherence mentioned above) are: (1) the number of elements in a schema and the number of schemata that belong to the same type; (2) distinguishability; (3) the level on which the schemata appear; (4) hierarchy.

1. The number of elements defines the range of directionality. Moreover, each range is dominated by an inverted U function, which points to an optimum of quantity. Chances of directionality decrease as we head for either extremity of this optimum. A larger range requires a larger number, up to a threshold set by psychoacoustic constraints. Outside the West there are, on the one hand, systems that consist of few elements (such as pentatonic scales) and non-discernible multi-element systems on the other. We shall discuss this further as we proceed.

2. Degree of distinguishability, or determining how similar or different things are. It relates to successive elements within a single schema (e.g., intervals in a specific scale) or between schemata of the same type (difference and similarity between modes or harmonic patterns). The more discrete they are the wider is the range of directionality and clarity of elements that are recognized by listeners because of immediate context. Outside the West there are many systems which contain non-discrete elements, especially rhythmic and inter-
valic, which are not defined categorically (Keefe et al. 1991). Preferring major and minor modes over all others in the Western modal system is one example of the increase in the intensity of categorization or the degree of distinguishability between schemata.

3. Level of occurrence: Complex schemata on an immediate level result in momentary directionality and complexity, yet make it difficult to maintain directionality on higher levels. This is known as Von Förster's law (Koppel et al. 1978), which presents it mathematically. Indeed, outside the West there exists an abundance of complex schemata on the immediate level, such as pre-determined rhythmic patterns which contain more than two types of beat and are also defined by timbre, an array of motifs, an abundance of scales which consist of more than two types of seconds and many inter-scale connections on the most immediate level, sometimes even connections to specific notes. In Western tonal music, on the other hand, we find only two types of second (major and minor), two types of beat (accented and unaccented), two types of meter (duple and triple) and two types of scale (major and minor) which can be handled through modulation and transformation to contrast highly directional and complex structures on various levels.

4. Hierarchy: Of course, any hierarchy leads to directionality. In the West in particular we notice strong hierarchy among scale-degrees, chords, tonal centers and so on. A given style may affect the enforcement of hierarchy. For example, we might draw a comparison between "typical Mozart" and "typical Bach" and consider the way in which chords schemata appear.

Natural schemata

Generally speaking, these schemata are depicted in the way various parameters are distributed across spaces of pitch and time, i.e., registers and contours of the various parameters (not only pitch) and also manipulations of learned and natural schemata. Contrary to learned schemata, natural ones can be represented already in the psychoacoustic stage and are not cultural conditioned. They cannot be defined in an exact quantifiable manner. Also, they are familiar to us from other sound phenomena which are non-musical (Cohen and Dubnov 1997). How do these phenomena fit into universal regularity which has already received many definitions (Gestalt laws, grouping into units, etc.)? What are the principles of natural (innate) schemata and their meaning?

It is worth noticing that theoretical considerations of expectation (Meyer 1956, 1973; Narmour 1989; Jones 1990) and experiments that tested listeners' responses (e.g., Dowling 1990; Yeger-Granot 1997) revealed innate regularity which assumes natural expectations, mostly of a relaxed tendency, as to the continuation of a given progression. Indeed, musical styles that obey an ideal of tension or excitement produce progressions that oppose either Gestalt law or Narmour's model.

Despite the numerous research studies on emotions in music, we would like to propose our own definitions of the main principles that cause a sense of tension and salience:

1. Intensified energy in every one of the parameters (higher, louder, denser, longer, larger ambitus). An example of this is the sense of accentuation on syllables caused by a peak in duration, pitch or loudness. They are interchangeable in respect to salience.

---

2 There even exist terms that reflect compositional rules that limit directionality on a higher level, such as "additivity" and "mosaic composition." This is yet another point for distinguishing between musical cultures. It prevails in 20th-century music, too.

3 Mozartian style obeys a stronger hierarchy than does that of Bach regarding the use of harmonic degrees, 7th chords, and secondary degrees.

4 From our point of view, the various kinds of natural schemata actually determine texture in its broader sense, which can be considered a super-parameter (Cohen and Dubnov 1997).

5 A fine example can be found in African songs which are based on tonal language, where at certain points a syllable, instead of being higher (pitch) becomes longer or accented and meaning is retained (Schneider 1957).
2. **A sudden change** in any of the parameters, such as sforzando, a large interval, and change of tempo. Palestrina counterpoint, which suggests an ideal of calmness, strictly forbids sudden changes. Apart from dissonants, there are other events that create tension, such as fast notes or large melodic intervals, which require "preparation" and "resolution" prior to and after their appearance, thus ensuring a certain graduation or rather the avoidance of sudden change.

3. **Inverted U function** means there exists a normative optimal size which regard the size of any occurrence. Any deviation to either one of the extremities (i.e., towards too little or too much less) will result in tension (Hargreaves 1986). This optimum is set for every parameter that may be set along a scale (not too high or low, not too fast or slow, etc.). So, tension is the result of either great change or none at all.6

4. **Nonconcurrence** of simultaneous contours of various parameters (e.g., when one rises as another declines; when salient events occur on unaccented beats; when the boundaries of simultaneous units do not concur; when the boundaries of learned and natural schemata do not concur). Nonconcurrence increases complexity and uncertainty and contributes to excitement. The factor of concurrence/nonconcurrence plays an important role in determining style (Cohen and Dubnov 1997).

5. **Curves resulting in uncertainty:** Six basic curves (convex, concave, zigzag, horizontal, ascending, descending) can be distinguished. We will mention focus on two of them here.

   Convexity means there is only one peak which follows a gradual rise and declines gradually. It seems to be an ideal contour for the prediction of how a progression would continue and serves as a unifying factor (we may also view this curve as reflecting symmetry), which might show up in various parameters on different levels.7

   The horizontal curve, or repetition, is of course necessary in any organization, but at least three different types can be noticed:

   a. **A single repetition of a directional unit**, such as the period in the classical era. A single such repetition heightens directionality. The same idea can be developed into a continuous repetition which keeps including previous units. This may be put as \( 2^n \) \((1 + 1 + 2 + 4 + \ldots)\) and is prominent in Far Eastern cultures, whereas in others it is absent. The West introduced \( 2^n \) in the 17th century along with establishing the major-minor, functional harmony, and tonality in general, i.e., those factors that make it possible to construct a complex overall structure.

   b. **An exact repetition of a small unit** increases tension up to a certain threshold where change is no longer anticipated. These repetitions were forbidden in Palestrina counterpoint, yet are quite common in the music of the Baroque.

   c. **Multiple repetition** which goes beyond that threshold helps perceive those events as background such as in a monotonous accompaniment which brings out main events. Bach’s homogeneous texture can easily become the main cause of interest because it constantly introduces latent yet meaningful changes.

5. **Deviation from expectations**, which Meyer considers to be the leading factor in creating tension, may be realized in two ways: by breaking learned schemata (as in harmonic progressions) or innate ones (convex curve or \( 2^n \)) or by occurring at the extremes of an inverted U function.

---

6 ERP experiments (Cohen and Granot 1993) have demonstrated strong reaction to both dissonances (no blend) and parallel fifths (too great a blend) whereas the interval of the third, as well as parallel thirds (optimal blend) have received minimal reaction. Again, Palestrina rules cling to an optimum in many respects.

7 It is no wonder that most folk tunes follow the pattern of convexity (Nettl 1968; Huron 1996). Palestrina counterpoint consists of convex curves on the immediate and phrase levels, for both pitch and duration. It has recently been found (Gurletsky 1995) that the overall structure in palestrinian masses is also convex in some respect. In musical styles meant to express excitement we shall find concave or zigzag curves.

608
We should emphasize that three levels of expectations can be detected (Cohen and Gra not 1995): the neutral level, on which we normally expect no change; the experience level, on which change is expected and all the above rules of tension are fulfilled; and the level of the stylistic ideal, on which expectations are guided by the laws of a given style and may stand in contradiction to "natural expectations."

Transformations

By transformations we mean repetitions that go through changes which are operation-defined. These transformations consist of the two basic elements of every organization: repetition and variance. They regard compositional rules and can be implemented on different parameters (within limitations that arise from those parameters) in musical events at the immediate level or any other schema, learned or natural at various levels.

We consider the transformations to be natural schemata since their principles can be viewed as innate cognitive processes. Generally speaking, we propose five principles of operations (Appel 1993): (1) contrast (elementary or complementary or binary); (2) augmentation and diminution; (3) shifting through a cyclic system; (4) fusion and segregation; (5) use of equivalence, which is a basic principle of any living language.

Transformations uphold intelligibility, but we may distinguish between degrees of intelligibility or directionality. The type of transformation (not described here in detail) is determined according to stylistic ideal.

Complex transformations have characterized Western music since the 17th century. They enable complex structures and directionality on various levels (together with other 17th-century schemata: 2^n, major and minor scales, functional harmony, the connection between melody and harmony dominated by a single tonality).

Outside the West the transformations are very limited: The operation of contrast is mostly applied gradually (not suddenly) on natural schemata in a rather simple manner — low-high, slow-fast, melismatic-syllabic, nonrhythmic-rhythmic, etc. Moreover, in many cases these gradual contrasts constitute an important part of the form of the piece. (In the West, in contrast, there are no predetermined restrictions on these contrasts, and therefore the composer has more freedom.) The operation of shifting is also limited. Augmentation and diminution hardly exist. Equivalence is dominant in all forms of music based on "modal frameworks" in which the "melody type" and improvisation are of great importance.

To sum up, natural schemata are revealed by the contours and range of distribution — taking into account the basic principles of tension and relaxation and the type of directionality (the degree of certainty regarding the continuation of the progression and the magnitude of the directional unit) — as well as by the five types of operations that can be considered cognitive operations.

Natural and Learned Schemata as Complementary Contrasts

If we accept the three conditions for the existence of a relationship of complementary contrast between two things — a contradiction between the two, with each one hindering the other; support; and identification with the function — then we see that the natural and learned schemata, which are contradictory in so many ways, are in a sense also complementary contrasts: (1) they may compete with each other for attention and even blur each other (in the case of nonconcurrence); (2) in situations of concurrence between them they may support and accentuate each other, as occurs prominently in the Classical period (Lorince 1966); (3) on a certain level the learned schemata may be regarded as being obtained from natural activity (a sort of cognitive "solution" to the demand of the stylistic ideal), and the natural schemata that have been shaped by people can be regarded as learned schemata.
An Analogy to Complementary Contrast: The Lexical and Prosodic Layers in Speech

Just as tonal music is based on learned schemata that are realized through various natural schemata, speech, as is well known, is based on a lexical layer that is realized through various manifestations of the prosodic layer. The lexical layer expresses the "what," i.e., the defined code of speech, whereas the prosodic layer, in addition to having a syntactical function, expresses the "how." In other words, the information it contains conveys the emotional message about the subject.8 The analogy is thus as follows:

<table>
<thead>
<tr>
<th>Lexical Layer</th>
<th>Learned Schemata</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prosodic Layer</td>
<td>Natural Schemata</td>
</tr>
</tbody>
</table>

From the acoustic standpoint the lexical layer relates to the timbres of phonemes, which are well defined in clear categories, just like the well-defined learned intervals.9 The prosodic layer, in contrast, is represented by curves of pitch, duration, and intensity that cannot be expressed in exact quantifiable manner; thus they are natural schemata. The emotion in the prosodic layer is expressed by the excitement factors mentioned above (nonconcurrence; deviations from the convex curve; adherence to the extremes of the U function with respect to tempo, register, degree of change, etc.). Another interesting analogy is between bird calls (not the "songs" of songbirds) and the prosodic layer in speech. A study of bird calls in nature (Cohen 1983; Katsir 1992) found that they convey information on the state of the bird (tense or calm) by means of the aforementioned excitement factors. This analogy underscores the universal validity of these factors.

Conclusion

We have seen that the natural and learned schemata may be classified on the basis of the meaning attached to them. The meanings are thus inherent in the schemata themselves, in the levels on which they appear, in the ways they are realized (including the deviations), and in the interrelations between simultaneous schemata. Learned schemata always appear with natural ones, whereas the natural schemata may appear even without the learned ones, as we find in 20th-century music, in which much of the notation involves natural schemata. In this case, some degree of freedom may exist, such that the same natural schema may be realized in various intervals and/or rhythms.

The theory of Western tonal music relates almost exclusively to clear learned schemata, which are suitable to complex overarching organization with directionality on various levels, while most natural schemata are not represented at all in the theory and are not involved in the definition of the form. They vary in accordance with the ideal of the period and the composer (variation in history is a characteristic of the West). In contrast, outside the West the learned schemata do not fulfill the conditions for coherence, and the natural schemata are used consciously, together with the learned ones, to define the musical form (thereby limiting the possibilities of stylistic changes).

In our opinion, the difference between the styles is not due to conventions regarding musical rules but rather to different aesthetic ideals that result from cultural convictions, social needs, or environmental conditions. The universality is manifested in the connection between the ideal and the selected schemata, which have potential for certain meanings and thus prevent even the learned schemata from being arbitrary.

Some of the ideas expressed here (very briefly) are based in part on experiments and analyses of various styles. Nevertheless, it is clear to us that to complete the picture additional tests must be carried out by means of experiments involving subjects from various cultures and comparative analysis of various kinds of music.

8Of the many research studies, we shall mention only those of Fónagy and Magdis (1972) and Sundberg (1982).

9We should stress, however, that this analogy between the phonemes and intervals (Rakowski 1990; Halperin 1986) is very limited and does not apply in other respects.
Music Experience
Experimental Analysis of the Perception of Sequences of Tones

R.J. Watt & R.L. Ash
Department of Psychology
University of Stirling
Stirling, Scotland FK9 4LA

There has been a longstanding tradition that music can be understood to have certain structural properties that allow it to be regarded as a system capable of communication (eg. Meyer, 1956; Sloboda, 1991). A listener to, say the 6th Symphony of Vaughan Williams, will have a definite and apparently spontaneous sense that something has happened by the end, although what that something might be is ambiguous.

The implication of the communication idea is that music somehow has a quasi-linguistic structure (eg Bernstein, 1976). Thus, it has been argued quite persuasively that music conforms to a generative grammar (Lerdahl and Jackendoff, 1983). That grammar, intended to represent the response of an ideal listener, has some formal similarities to the grammar of natural language: it handles recursive structures as such, it is not equivalent to phase-transition automata, nor to phrase-structure grammars. More contentious is the question of whether that grammar is applied to a vocabulary (implying semantic elements of structure).

The clearest exposition of this proposition is by Cooke(1959) who explicitly identified a primitive vocabulary for western tonal music. This vocabulary has two distinct properties that lend credence to the notion that it represents an elementary referential system. First, the members of the vocabulary, which are short sequences of notes (generally between 3 and 6 pitches or notes) are outstandingly common sequences in western music. If one takes all possible sequences of 3 to 6 pitches (disregarding octave distinctions), of which there are 19600 using just the diatonic scale and 177144 using the full chromatic scale, then a tiny number of these are extremely common. Cooke identifies these as the vocabulary of music. The second property that he identifies is a consistent ‘meaning’ for each of these common sequences. The meaning is expressible in loose terms as being joyful or triumphant or whatever, but appears in his analysis to apply irrespective of the date, nationality or context of the music.

One can place alongside this style of analysis another aspect of music. At any one period of history (at least for western tonal music), there are certain configurations of notes that are regarded, technically speaking, as being unstable. Such instabilities are required, by technical convention to move into stability. Dissonances are required to resolve into consonance. That these are essentially technical is clear from the manner that the set of pitch configurations that are regarded as being dissonant changes with the age. Thus at one time, only the perfect intervals were considered consonant, with major thirds treated as dissonances. Later, thirds and sixths became consonant leaving the seconds and the augmented fourth as the only dissonances. More recently, the latter interval has acquired much of the status of a consonant interval (Britten’s music frequently uses it as such).
That the concept of consonance and dissonance does not have a fixed mapping on to music, but depends on musical context appears to render any simple notion of a vocabulary for music to seem untenable. This is not quite correct however as the following argument will illustrate.

Start by supposing that there is some form of natural perceptual effect of note pairs (simultaneous for the sake of argument), just as Helmholtz suggested (1885). So:
- octaves are the purest;
- perfect intervals the next simplest;
- major third, minor sixths more complex;
- minor thirds, major sixths follow next;
- and so on.

Then suppose that this ordering is fixed and represents the basic building bricks with which music is perceived. So, whatever the type of music (within limits) the first cognitive representation is formed of values along this ordering. A simple step towards a less cognitive and a more affective response can be made by supposing that movement down the list causes a increase in tension and movement up the list, relaxation. This is not quite right because it implies that there is never any direct sense of location within the list, so a major third would be consonant if it followed a major sixth, but dissonant if it followed a perfect fifth. This does not accord with experience.

Suppose that subjects are aware of location on that list, at any one time, but relate their location not to just the immediate preceding event, but their recent history (or even their entire lifetime’s experience). Thus if the location they are hearing at the moment is on the low edge of their experience of the list, then that is sensed as tense, whereas if they are hearing an event located well within normal experience, it is not tense.

This is type of account would be described as an error-correcting perceptual mechanism (eg. Andrews, 1964) and is not a very novel idea. It has been spelled out because it requires two experimentally verifiable suppositions. First, it is supposed that there is a recognition of position on the list above (and some basically equivalent lists). Second, it is supposed that this is different from the effective perception of a piece of music in a very general sense. Therefore, it is necessary to show that ordinary listeners, expert in music through relatively passive experience rather than active education, are sensitive to the gradations implied by the list structure. Next it is necessary to show that listeners perceptions of the basic interval level structure is qualitatively different from their perception of whole pieces of music.

In a previous experiment, we explored the question of the perception of (relatively) whole pieces of music. We played pieces of music to subjects and then had them make forced choices from pairs of approximately opposite words. Pairs of words were chosen to represent:
- descriptors of people (such as male/female);
- emotions or moods (such as happy/sad);
- spatial terms (such as narrow/wide);
- terms that can be used in fairly general domains (such as sweet/sour);
- control pairs that are not obviously applicable (such as yellow/violet).

After hearing each piece of music in turn, subjects were required to choose one member from each of a set of 14 pairs belonging to the above categories. The main finding from the experiments was that the subjects showed highest agreement for their choices between pairs of terms when the terms could be applied to people; the next highest agreement levels were for
emotional and mood terms. These categories could generate levels of agreement of the order of 90% across 200+ subjects.

The implication of this finding is that there exist dimensions along which music is easily discriminated that can be mapped reliably onto dimensions that serve to distinguish people. A short-hand account of the finding is that the subjects can map variations in style and nature of music onto variations in people. At one level this is not surprising: we are experts in the latter part of the proposition, knowing a great deal about variations in people, and it may be that we can map nearly any stimulus onto that set of variations.

For this reason, we repeated the experiment with another set of stimuli, sharing a degree of sensory complexity and personal preference. The set of stimuli we chose were pieces of food. Subjects ate a piece of food and then performed the same forced-choice task. Quite generally, there was little agreement between subjects, with all categories showing less than 65% consistency of responses across the population. Thus the finding for music is not simply a general response to any stimulus, but instead represents some degree of expertise in music and some form of natural mapping from music to representations of people.

As a follow on to those experiments, we have addressed the question of the perception of much shorter fragments of music, using the same technique and now report the results. Two specific questions are asked:

1. How long does a sequence of notes need to last for it to have the full perceptual effect of music?
2. What is the nature of perceptual responses to shorter sequences of music?

In each case the subjects used were not formally trained in music, and had all acquired their sensitivities to music through normal listening.

**Experiment 1**

In the first experiment, the responses of subjects to segments of three different gaelic melodies were measured, as a function of segment length. Subjects heard a stimulus which was an extract from one of the three melodies, and were then required to make a set of forced-choice responses to category pairs as described above. The level of agreement between subjects' responses to person-related response categories (male/female, good/evil, pleased/angry, gentle/violent) was compared with the level of agreement on non-person-related categories.

Figure 1 shows the effect of the extract duration (in beats) on the degree of agreement for these two different sets of response categories for the three melodies. As can be seen there is no effect of duration on the non-person-related categories, which remain uniformly low. There is a large effect of duration on the person-related categories. These start at the same level of agreement as the non-person-related categories, but after 3 beats show a rather sudden and large increase in agreement. From this point onwards there exists a statistically significant difference between the two sets of responses (person and non-person).

**Experiment 2**

In the second experiment, the same procedure was applied to sequences of two notes. 48 such sequences, constrained to be within a single octave C4 to C5, were used: the 12 intervals from
C4 upwards; the 12 intervals from C5 downwards; the 12 intervals down to C4; and the 12 intervals up to C5. The results obtained generated relatively low degrees of agreement across subjects (although still statistically significant). Of interest is the underlying basis for such responses. This was revealed by performing a Principal Components Analysis on the data, and projecting the first few factors onto the stimulus set. If this projection has a ready interpretation, then that would indicate the likely psychological basis for the responses.

Figure 2 shows the first two factors projected back onto the set of interval sizes. As can be seen, the first factor has a linear trend with interval size, perturbed by the intervals involving non-diatonic notes. These latter are accorded a response as if they were perceived with the non-diatonic notes being higher in pitch than they actually are. The second factor corresponds more simply to a division of the intervals into those that are a third or less and those that are not. As can be seen from the histogram of interval distribution in the whole set of 32 gaelic melodies, this corresponds quite closely with a distinction between melodic interval sizes and non-melodic intervals.

**Conclusion**

The experimental results have shown that subjects are indeed sensitive to immediate diatonic interval structures within music. This is not surprising, although it is in contradiction to some earlier studies that have found that musically untrained listeners do not show a sensitivity to the distinction between diatonic and non-diatonic structures (Krumhansl 1979). The results also show a radical change in perception of note sequences after around 3 beats worth of music. Up until that point, interval structures presumably determine judgments. After 3 beats, subjects switch to a different form of perception where the music can generate reliable person-related responses.

These two results appear to lend some support to the contention that certain sequences of notes function as a vocabulary in the sense that they cannot be subdivided without a loss of meaning.

**References**


Helmholtz HLF. (1885/1954) On the Sensations of Tone. (trans A. Ellis). Dover


Information in subjects’ responses to pieces of music as a function of musical extract duration for non-person categories of response and person categories of response. The latter show a large increase in information after about 3 beats of music. Information is defined as percentage agreement (50% to 100%) scales to the range 0 to 1. Thus 0 information corresponds to 50% agreement (i.e., chance) and 1 information corresponds to 100% agreement.
Figure 2

This figure shows the first two components of subjects responses to sequential intervals. The first component, which accounts for most of the variance, corresponds to interval size, subject to a transformation which lends a greater apparent interval size to intervals involving non-diatonic notes. The second component corresponds to a distinction between melodic and non-melodic intervals. The dashed curve shows the actual distribution of intervals in a larger set of gaelic melodies.
Effect of Video Presentation on Asian Music Perceptual Dimensions
C. Victor Fung
Hong Kong Baptist University

The purpose of this study was to determine the effect of video presentation on perceptual dimensions used by students in classifying Asian musics and on preferences for Asian musics. Specific research questions were: (1) What were the Asian music perceptual dimensions in video presentations? (2) What were the Asian music perceptual dimensions in audio presentations? (3) How were Asian music preferences related to audio stimulus characteristics and video stimulus characteristics?

Method

Instrument.
Asian musics comprised 15 musical excerpts selected from the first 15 volumes of the JVC Video Anthology of World Music and Dance. The complete video anthology consists of 30 volumes and the first 15 volumes include only Asian musics. The anthology is one of the most comprehensive of its kind but omits Japanese music. Reviewers of this anthology (Anderson, 1991; Carty, 1991; Miller, 1993; Wong, 1995) suggested that it could enrich the classroom music experience tremendously if the instructor used it with great care.

A music educator with training in ethnomusicology categorized all 164 excerpts in the first 15 volumes of Asian video into: (1) stage settings (64 excerpts)--showing musicians (not dancers or actors) performing in stage settings, (2) dance and drama (61 excerpts)--showing dancers and actors/actresses as the focus with music at the background, and (3) field settings (39 excerpts)--showing musical activities as rituals, festivals, and social events other than performances in stage settings. Using a table of random numbers, five excerpts were selected from each of the three excerpt categories (See Table 1).

Table 1
Asian Titles Randomly Selected from the JVC Anthology of World Music and Dance

<table>
<thead>
<tr>
<th>Series number and title</th>
<th>Region*</th>
<th>Setting**</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-3 Salp'uri--dance to exorcise evil spirits, Korea</td>
<td>EA</td>
<td>D</td>
</tr>
<tr>
<td>3-4 Zheng--zither solo: &quot;Yuzhou changwan,&quot; Han, China</td>
<td>EA</td>
<td>S</td>
</tr>
<tr>
<td>4-1 Dage--group dance, Yunnan, China</td>
<td>EA</td>
<td>F</td>
</tr>
<tr>
<td>5-2 Lullaby, Xinjiang, China</td>
<td>EA</td>
<td>F</td>
</tr>
<tr>
<td>5-5 Song with dutar and daf, Xinjiang, China</td>
<td>EA</td>
<td>F</td>
</tr>
<tr>
<td>5-6 Panjgah-mukam--classical music, Xinjiang, China</td>
<td>EA</td>
<td>S</td>
</tr>
<tr>
<td>7-11 Monkey play, Thailand</td>
<td>SEA</td>
<td>D</td>
</tr>
<tr>
<td>8-11 Tongatong--stamping tubes, Philippines</td>
<td>SEA</td>
<td>S</td>
</tr>
<tr>
<td>8-12 Topayya--gongs struck with the hands, Philippines</td>
<td>SEA</td>
<td>S</td>
</tr>
<tr>
<td>9-4 Legong keraton--court dance, Bali, Indonesia</td>
<td>SEA</td>
<td>D</td>
</tr>
<tr>
<td>11-7 Kathak--classical dance: &quot;Tarana,&quot; India</td>
<td>SA</td>
<td>D</td>
</tr>
<tr>
<td>12-5 Bhajan--Sikh devotional song, India</td>
<td>SA</td>
<td>F</td>
</tr>
<tr>
<td>13-1 Rudra-vina--plucked zither solo, India</td>
<td>SA</td>
<td>S</td>
</tr>
<tr>
<td>14-6 Tarwar--sword dance, Pakistan</td>
<td>SA</td>
<td>F</td>
</tr>
<tr>
<td>15-4 Dance from Perahara festival: &quot;Ves.&quot; Sri Lanka</td>
<td>SA</td>
<td>D</td>
</tr>
</tbody>
</table>

*EA=East Asia, SEA=Southeast Asia, SA=South Asia.
**D=Dance and drama, F=Field settings, S=Stage settings of musicians.
Only the last 15 seconds of each selected excerpt showing the most comprehensive views of the musical activities were used in this study. That is, segments with distractions such as video shots on the scenery of the country, video shots that included subtitles or headings, or video shots that focused on the spectators rather than the performers were not used.

Four stimulus tapes were created using the selected excerpts. All stimulus tapes presented 15 excerpts in all 105 possible pairs for comparison. Two tapes had both video and audio tracks. Two tapes had the audio track only (i.e., sound on video tape without the video image). Excerpts were recorded in random order. To reduce the order effect in pair comparisons, excerpt pairs in one video-audio tape and one audio-only tape were recorded in reverse order.

Subjects.
Subjects were 49 undergraduate nonmusic majors enrolled in a general arts course at a large public university in the U.S.A. There were 22 females and 27 males. Subjects' ages ranged from 18 to 30 years (mean=20.38, median=20, SD=2.45). The sample of subjects in this study was somewhat diverse, consisting of 51.0% Caucasian-American, 26.5% Asian-American, 12.2% African-American, 2.0% American-Indian, 2.0% Latin-American, and 6.1% other.

Procedure.
Subjects were randomly assigned to receive musical stimuli from one of the four video tapes. Twenty-seven subjects responded to video-audio tapes and twenty-two subjects responded to audio-only tapes. The uneven number of subjects in each group was due to subject attrition. Subjects were asked to provide similarity ratings (1=highly dissimilar, 7=highly similar) for all 105 possible pairs of musical excerpts. Students also indicated which of the excerpts was preferred/liked in each pair. In addition, two other pairs of Asian musical excerpts were used at the beginning of each stimulus-tape for participants to practice responding to the stimuli, and they were not used for analysis. Subjects responded to the excerpt pairs individually. They completed responding to stimulus-pairs in about 75 minutes including a short break in the middle.

In addition, two ethnomusicologists (members of the Society for Ethnomusicology with graduate degrees) viewed the same stimuli and provided musical characteristics ratings. These ratings were based on nine musical characteristic rating scales used for world musics (Fung, 1996): Slow/Fast, Redundant pitches/Different pitches, Non-tonal centered/Tonal centered, Dissonant/Consonant, Dull timbre/Bright timbre, Smooth/Percussive, Soft/Loud, Simple texture/Complex texture, and No embellishment/Rich embellishment. Four rating scales were added for this study: Dominated by voice(s)/Dominated by instrument(s), Difficult to dance with/Easy to dance with, Little visible movements/Active visible movements, and Dull visual colors/Bright visual colors. All scales used seven-points, with the first characteristic as one and the second characteristic as seven.

Results
Data were analysed using the SPSS. Three scales with low interjudge reliability coefficients (r<.70) were not included in subsequent analyses (Non-tonal centered/Tonal centered, Dissonant/Consonant, and No embellishment/Rich embellishment). The remaining ten scales had reliability coefficients ranging from .73 to .96. Experts' ratings on stimulus characteristics are presented in Table 2. Subjects' preferences for each excerpt based on pair comparisons were also presented in Table 2.
Table 2
Preference Ratings of the Excerpts Using Paired Comparison Techniques by Undergraduate Students and Ratings of the Excerpts by Experts on Ten Characteristic Scales

<table>
<thead>
<tr>
<th>Excerpt</th>
<th>Preference</th>
<th>Stimulus Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Video</td>
<td>Audio</td>
</tr>
<tr>
<td>Salp’uri</td>
<td>7.111</td>
<td>8.409</td>
</tr>
<tr>
<td>Zheng</td>
<td>10.111</td>
<td>10.345</td>
</tr>
<tr>
<td>Dage</td>
<td>7.185</td>
<td>6.136</td>
</tr>
<tr>
<td>Lullaby</td>
<td>4.370</td>
<td>3.000</td>
</tr>
<tr>
<td>Song</td>
<td>6.111</td>
<td>5.909</td>
</tr>
<tr>
<td>Panjgha</td>
<td>6.778</td>
<td>6.773</td>
</tr>
<tr>
<td>Monkey play</td>
<td>6.185</td>
<td>5.591</td>
</tr>
<tr>
<td>Tongatong</td>
<td>5.630</td>
<td>6.136</td>
</tr>
<tr>
<td>Topayya</td>
<td>6.333</td>
<td>6.409</td>
</tr>
<tr>
<td>Legong</td>
<td>9.074</td>
<td>8.318</td>
</tr>
<tr>
<td>Kathak</td>
<td>7.407</td>
<td>6.955</td>
</tr>
<tr>
<td>Bhajan</td>
<td>4.333</td>
<td>5.045</td>
</tr>
<tr>
<td>Rudra-vina</td>
<td>7.333</td>
<td>8.364</td>
</tr>
<tr>
<td>Tarwar</td>
<td>7.778</td>
<td>8.091</td>
</tr>
<tr>
<td>Dance</td>
<td>9.407</td>
<td>9.182</td>
</tr>
</tbody>
</table>

Fast=Slow-Fast, Diff=Redundant pitches-Different pitches, Brig=Dull timbre-Bright timbre, Perc=Smooth-Percussive, Loud=Soft-Loud, Simple texture-Complex texture, Inst=Dominated by voice(s)-Dominated by instrument(s), Dance=Difficult to dance with-Easy to dance with, Move=Little visible movements-Active visible movements, Colo=Dull visual colors-Bright visual colors

Pearson Product-Moment correlation was used to examine the relationships between preference and stimulus characteristics. Instrumental dominance (r=.70 and .82), fast tempo (r=.69 and .63), and bright timbre (r=.51 and .52) were significantly correlated with preferences presented in both video and audio formats. Loud sounding (r=.57), active visible movements (r=.56), and bright visual colors (r=.58) were significantly correlated with preferences presented in video format only.

Two separate matrices of proximities, one for subjects receiving video stimuli and one for subjects receiving audio stimuli, were obtained between all 15 musical excerpts. The matrices were analysed using separate multidimensional scaling analyses (ALSCAL). A three dimensional solution was found for each group. These solutions were found to account for greater share of variance in similarity ratings ($r^2=.90$, Kruskal’s stress=.11 for the video group; $r^2=.83$, Kruskal’s stress=.14 for the audio group) when compared to the two-dimensional solution ($r^2=.81$, Kruskal’s stress=.18 for the video group; $r^2=.78$, Kruskal’s stress=.20 for the audio group). The three-dimensional solutions were adopted for further consideration.

With reference to the stimulus configurations for the video group and the audio group, there were some similarities and differences between the groups. The video group shows more or less 4 neighborhoods of the space. *Lullaby* stands out alone in space featuring solo voice, field setting, extreme textural simplicity, and not easily danceable character. *Tongatong* and *Topayya* stand out at the opposite direction of *Lullaby* featuring highly redundant pitches, percussive sounds, and instruments alone with stage settings. The others were in between these two neighborhoods with a mixture of characteristics. The audio group shows less clear-cut neighborhoods. However, three excerpts stand out at the greatest distances to each other: (1) *Lullaby*, the least preferred of all audio excerpts, (2) *Zheng* featuring a string instrument solo with freely changing tempo, and bright timbre, and it was the most preferred audio excerpt, and (3) *Tongatong* featuring loud, percussive, and repetitive sounds.
Correlation of the three dimensions between the video and audio groups was sought through a series of six multiple regression analyses in which each of the dimension coordinates of each dimension were regressed on the coordinates of all dimensions of the other group. Each dimension in each group served as the dependent variable in each of the six regression analyses. Results are shown in Table 3. The first dimension of both groups shared 83.83% of variance. Dimension 3 of the video group and Dimension 2 of the audio group shared 61.39% of variance. However, both Dimension 2 of the video group and Dimension 3 of the audio group were unique in that they did not share significant variance with any dimension of the other group. Therefore, data from the two groups were not combined in subsequent analyses.

Table 3
Stepwise Multiple Regression Summaries with Various Dimensions as Criterion Variables and Predictor Variables

<table>
<thead>
<tr>
<th>Predictor</th>
<th>MR</th>
<th>r</th>
<th>R²</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video Dimension 1 as criterion variable:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio Dim. 1</td>
<td>.9156</td>
<td>.9156</td>
<td>.8383</td>
<td>67.37</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Video Dimension 2 as criterion variable:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No audio dimension was entered</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Video Dimension 3 as criterion variable:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio Dim. 2</td>
<td>.7836</td>
<td>.7836</td>
<td>.6139</td>
<td>20.67</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Audio Dimension 1 as criterion variable:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Video Dim. 1</td>
<td>.9156</td>
<td>.9156</td>
<td>.8383</td>
<td>67.37</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Video Dim. 2</td>
<td>.9570</td>
<td>.2995</td>
<td>.9159</td>
<td>65.34</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Audio Dimension 2 as criterion variable:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Video Dim. 3</td>
<td>.7836</td>
<td>.7836</td>
<td>.6139</td>
<td>20.67</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Video Dim. 2</td>
<td>.8531</td>
<td>.3323</td>
<td>.7278</td>
<td>16.04</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Audio Dimension 3 as criterion variable:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No video dimension was entered</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Interpretation of the three dimensions was sought through another series of 11 multiple regression analyses for each group. Experts’ ratings on 10 stimulus characteristics and subjects’ preferences for excerpts were regressed on the stimulus coordinates. Each stimulus characteristic and preference served as the dependent variable with the coordinates of the configuration as independent variables. Results are shown in Table 4. Only the beta weights of significant predictors are listed. For the video group, all stimulus characteristics, except textural complexity, and preference were significantly correlated with the location of excerpts in the multidimensional scaling solution. Most of the stimulus characteristics were significantly correlated with coordinates of Dimension 1 only. Visual movement was related to both Dimension 1 and Dimension 2. Danceability and preference were related to Dimensions 1 and 3. For the audio group, all stimulus characteristics, except timbre brightness and textural complexity, and preference were significantly correlated with the location of excerpts in the multidimensional scaling solution. All significant characteristics were correlated with coordinates of Dimension 1. In addition, three stimulus characteristics were related to both Dimension 1 and Dimension 2: percussiveness, domination of instrument or voice, and danceability. Preference was related to both Dimension 1 and Dimension 2 also. Dimension 3 was not related to any stimulus characteristics or preference.
Table 4
Multiple Regression Results Using Stimulus Characteristics and Preference as Criterion Variables and the Stimulus Coordinates of the Three Dimensional Solution as Predictor Variables

<table>
<thead>
<tr>
<th></th>
<th>$MR$</th>
<th>$R^2$</th>
<th>$F$</th>
<th>$p$</th>
<th>Beta 1</th>
<th>Beta 2</th>
<th>Beta 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video Group:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fast</td>
<td>.7531</td>
<td>.5671</td>
<td>17.031</td>
<td>&lt;.001</td>
<td>-.7531</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diff</td>
<td>.6942</td>
<td>.4820</td>
<td>12.095</td>
<td>&lt;.01</td>
<td>.6942</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brig</td>
<td>.5801</td>
<td>.3365</td>
<td>6.594</td>
<td>&lt;.05</td>
<td>-.5801</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perc</td>
<td>.8446</td>
<td>.7133</td>
<td>32.348</td>
<td>&lt;.001</td>
<td>-.8446</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loud</td>
<td>.8897</td>
<td>.7915</td>
<td>49.361</td>
<td>&lt;.001</td>
<td>-.8897</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comp No dimemsion was entered.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inst</td>
<td>.6624</td>
<td>.4387</td>
<td>10.162</td>
<td>&lt;.01</td>
<td>-.6624</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Danc</td>
<td>.8380</td>
<td>.7023</td>
<td>14.156</td>
<td>&lt;.001</td>
<td>-.6751</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Move</td>
<td>.9207</td>
<td>.8477</td>
<td>33.397</td>
<td>&lt;.001</td>
<td>-.7874</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colo</td>
<td>.7302</td>
<td>.5332</td>
<td>14.852</td>
<td>&lt;.01</td>
<td>-.7302</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preference</td>
<td>.8211</td>
<td>.6742</td>
<td>12.417</td>
<td>&lt;.001</td>
<td>-.5968</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio Group:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fast</td>
<td>.7420</td>
<td>.5505</td>
<td>15.920</td>
<td>&lt;.01</td>
<td>-.7420</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diff</td>
<td>.7478</td>
<td>.5592</td>
<td>16.489</td>
<td>&lt;.001</td>
<td>-.7478</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brig No dimemsion was entered.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perc</td>
<td>.9440</td>
<td>.8912</td>
<td>49.139</td>
<td>&lt;.001</td>
<td>-.9075</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loud</td>
<td>.8438</td>
<td>.7120</td>
<td>32.137</td>
<td>&lt;.001</td>
<td>-.8438</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comp No dimemsion was entered.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inst</td>
<td>.8951</td>
<td>.8013</td>
<td>24.191</td>
<td>&lt;.001</td>
<td>-.8117</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Danc</td>
<td>.7782</td>
<td>.6056</td>
<td>9.214</td>
<td>&lt;.01</td>
<td>-.5712</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Move</td>
<td>.5885</td>
<td>.3464</td>
<td>6.889</td>
<td>&lt;.05</td>
<td>-.5885</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colo</td>
<td>.5602</td>
<td>.3138</td>
<td>5.946</td>
<td>&lt;.05</td>
<td>-.5602</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preference</td>
<td>.8504</td>
<td>.7231</td>
<td>15.669</td>
<td>&lt;.001</td>
<td>-.5556</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fast=Slow-Fast, Diff=Redundant pitches-Different pitches, Brig=Dull timbre-Bright timbre, Perc=Smooth-Percussive, Loud=Soft-Loud, Simple texture-Complex texture, Inst=Dominated by voice(s)-Dominated by instrument(s), Danc=Difficult to dance with-Easy to dance with, Move=Little visible movements-Active visible movements, Colo=Dull visual colors-Bright visual colors

To help interpret the perceptual dimensions in relation to geographic region (East Asia, Southeast Asia, and South Asia) and setting (dance and drama, field settings, and stage settings), two separate MANOVA's (one for video group and one for audio group) were computed for each group using the coordinates of dimensions as dependent variables. Results show that setting had a significant effect for the video group ($p<.05$). All other main effects (region for video group and audio group, and setting for audio group) were not significant. Univariate $F$ tests show that Dimension 2 of the video group had a significant difference depending on the setting. Setting explained 58% of the variance in Dimension 2 of the video group.

Conclusions

Preference for audio excerpts was related to instrumental dominance, fast tempo, and bright timbre. Preference for video excerpts was related to the same characteristics, and in addition, it was related to loudness, and two visual qualities (active movements and bright visual colors).
Results of this study suggest that video stimuli and audio stimuli share some commonalities in students’ perceptions, and at the same time, each type of stimulus material (audio and visual) shows a unique perceptual dimension. This justifies the value of using both audio and video materials for music instruction.

The first perceptual dimension of both groups (audio and video) was highly similar, sharing about 84% of the variance. This dimension was explained by a range of musical and visual characteristics and preference.

The second dimension of the audio group and the third dimension of the video group shared about 61% of the variance. Both of these dimensions, although from different groups, were explained by preference and danceability.

The second dimension of the video group was a unique dimension, unrelated to any dimension of the audio group. It was explained by active visible movement. In addition, it was also related to the setting of video excerpts (dance and drama, field, or stage). This dimension contributed to the visual contexts of the stimuli and was considered crucial to many ethnomusicologists. Spatial relationships and contextual settings could well be integral parts of music that were not captured in audio presentation alone.

The third dimension of the audio group was also a unique dimension, unrelated to any dimension of the video group. However, it was undetermined in this study. Musical or visual characteristics, preference, setting, and geographic region did not explain this dimension.

Probably the most surprising finding of this study was that geographic region did not help interpret any of the perceptual dimensions. Despite the fact that most teachers and textbook writers of world musics used the geographic approach to organize their curricula (Fung, 1995), students did not seem to perceive excerpts in relation to the geographic origin of the music.

There are still plenty of unanswered questions in the area of music perception using videos and using world musics. With the continuous growth of innovative music presentation formats, such as CD-ROM, karaoke, and virtual reality, more research is warranted.

References
Social-Perceptual Mechanisms of Musical Experience

L.L. Bochkarev
Russian Institute for Cultural Research, Moscow, Russia

Listeners' perceptions of external, expressive reactions of performers significantly affects the process of comprehension of the composer's conception and the performer's interpretation. That is, musical perception in immediate (concert) and intermediate (television, cinema) conditions differs from listening to music on the radio (recorded). The performer's expression, his gestures, behaviour, and external emotional expression can significantly affect the process of musical communication. Telecasting permits a good picture of the conductor's hands and the performer's face.

The possibility of feedback of this kind of perception is the possibility to switch the centre of attention from aural to visual. It is a well-known fact that we are able to perceive more information via vision than via audition. Some experts (T. Adorno, G. Anders, A. Dilijan, B. Trowel) stick to the opinion that music should not be telecasted because the visual sequence prevents adequate aural perception. The results of the discussions are summed up in the works of Korykhalova and Kozandzhieva-Velinova.

The aim of this study, the results of which are partly presented in the present paper, is an experimental investigation of the degree of influence of externally expressed forms or behaviours of a performer of music on the adequacy of cognition of emotional images of music under conditions of intermediate (telecasting) communication.

Methods

With the help of video technique, a conflict between aural (musical) and visual (expressive) information was created. This permitted an examination of the level of emotional significance of the visual program, and the degree of its contribution to the formation of the musical image.

Short musical pieces written by H. Asakawa, B. Bartoc, D. Dowland, and F. Chopin on definite topics were used. In the present paper, the results of the perception of the Prelude in E minor (No.4, op.28) and the exposition of Nocturn in C minor (No.13, op.48) by Chopin are presented.

In both pieces, two contrasting emotional states, "tenderness" and "despair", are expressed. The exposition of the Nocturn is lyrical, tender, contemplative, full with sweet sorrow. The melody is written in a high register backed by basses of sorrow, despair and suffering. Disquieting accompaniment fills the contemplative, languish melody of the Prelude with the feeling of despair.

The performers that took part in the experiment - professor A and student B from the Conservatory - listened several times to the interpretation, and after this they performed the music on piano backed by soundtrack (synchronized with the sound on the record). In the first case, they stressed external expression of positive emotion on their faces and sweet and lyrical feeling of the musical message (variant "Tenderness"), and in the second case they stressed external expression of negative emotions: sorrow, despair, mourning (variant "Despair"). In both variants ("T" and "D") external expressive
reactions of the performers were connected with different individual perception of the music. According to the task set by the experimenter, they should not contradict the conception of the composer or disrupt the continuity of the musical image, but they should be contrasting (generalized titles of the variants conveniently expressed polar expressive concepts).

Emotional reactions and expressions of the performers were video recorded with the soundtracks of the interpretations by eminent performers. High accuracy of synchronization permitted integration of the visual and the aural medium.

In Experiment I, listeners (2nd year students from all departments of the Conservatory and teachers from musical schools, 96 in total) evaluated the dynamics of the external expressions recorded on video without the soundtrack. They used a scale with a set consisting of 50 names of emotions and psychic states. From the proposed list they had to choose 2–4 definitions reflecting expressive content of the musical message and to draw a ring around the dominating emotion. They evaluated the degree of aesthetic attraction and expression of the performers and the quality of performing.

In Experiment II, one and the same recording of the Nocturn (performed by M. Vookresenskaya) and the Prelude (performed by N. Akopjan) were presented to the listeners in turns with different variants of emotional expression of the performers ("T" and "D"). The listeners evaluated the emotional content of the interpretations of the music according to the technique given above. In Experiment III, different interpretations of the music were shown with one video recording. In Experiment IV, they listened to the music without video support and they had to compare the interpretation with earlier variants and to name the performer.

Results

Both performers succeeded in meeting the requirements of the experimenter; to show mainly positive emotions in the variant "Tenderness", and predominately negative emotions in the variant "Despair".

During perception of performer B's expression without the soundtrack (variant "T"), 53,2% of the listeners saw tenderness, 38,3% - sweetness, 21,3% - bliss, 14,9% - high spirit, and only 4,2% sorrow, 8,5% melancholy. In evaluation of the "D" variant, none of the listeners named any positive emotions, 24,3% of the listeners have saw sorrow, 18,9% - depression, 13,5% - regret, 24,3% intellectual efforts, 16,2% - suffering. Similar tendencies are reflected in the evaluations of performers A's expression.

The comparative analysis of the video and the performer's expression testified to the fact that the performer B had more intensive mimics, the listeners applied more names of emotions in the evaluation of his expression. The listeners put higher aesthetic marks to the expression of this performer: 86,7% evaluated expression as positive, whereas only 42,8% evaluated the expression of performer A as positive.

If in Experiment I the attention of the listeners was drawn to the expression of the performers only, in Experiment II it was divided between aural (main) and visual (auxiliary) information. In all "musical" variants compared with "non-musical" variants, the quantity of the selected types of negative emotions decreased dramatically, which
undoubtedly can be ascribed to the influence of the emotional message of the music in minor. Thus, during perception of musical variant "T" performed by B 29,3% of the listeners selected "melancholy", 19,1% - "sorrow", 19,2% - "regret", 17% - "grief", 4,2% - "suffering". During perception of the expression without soundtrack, the total percent of the negative emotions was only 12,7%. But in the musical variants, the evaluation of the listeners during perception of the variants "T" and "D" were dramatically contrasting, which can be connected with the influence of the external expression on the perception of the emotional message of the music. Names of positive emotions were dominating; "tenderness" - 38,2%, "high spirits" - 23,4%, "yearning" - 19,2%, "sweetness" - 17%, "agitation" - 12,8%, "bliss" - 6,2%, "joy" - 4,2%, "sorrow" - 57,1%, "grief" - 48,4%, "melancholy" - 28,6%, "regret" - 37,1%, and suffering - 12,3%. A similar result was detected during comparison of the selection of every emotion in percentile relation to the total number of all listed emotions (all deviations are statistically significant at p<0.01).

All of the conservatory students and 91,6% of the teachers stated that they heard another interpretation when they listened to one and the same soundtrack with different variants of expression, and they tried to find now non-existent features in it. Undoubtedly, it can be explained by the role of the frame that is not set intentionally by the experimenter.

The influence of the frame during perception of one and the same soundtrack with expressive visual medium of different performers is especially noticeable. The majority of tests "heard" non-existent nuances, faults in articulation, phrasing, and pedalling during perception of one and the same soundtrack with different variants of the expression.

During perception of one and the same variant of expression with different variants of interpretation (Experiment III) 70% of students and 55% of the teachers stated that they heard different interpretations, but they gave more similar emotional evaluations compared with Experiment II.

The absence of the video support optimized perception to an even greater degree: In Experiment IV, 95,6% of the teachers and 55,6% of the students from the department of musical theory, and 77,8% from the piano department, gave adequate evaluations of a completely new recording which was not used in the experiment previously.

The main part of the students in Experiment II and III liked the performance and expression by the fellow-student more than the performance by the teacher: 88,9% listeners gave positive markings to the performance and 78,1% to the expression of performer B. The percent of negative marks was only 4,4%. Only 57,4% of the students gave positive markings to the performance of the teacher, and 45,8% gave positive markings to his expression. Percent of negative markings fall respectively between 31,3 and 35,7. These data agree with the results of investigations in the field of social perception, in that people tend to assess more productively and to differentiate better expressions representative of their age group.

Comparison of the assessment of emotional message of the musical interpretations done by the performers of emotional and intellectual types testifies to the fact that the degree of intensity of external expression of emotions significantly affects the assessment of the emotional message of the music.
Conclusions

1) Attenuation technique used in the analysis permitted an experimental determination of the degree of emotional significance of the externally expressed forms of emotional expression of a musician–performer affecting the adequacy of perception of emotional message of the music by the listeners. In contrast to the experiments by I.E. Gersamee, Z.G. Kazanjieva–Velinova, W. Probst, M.G. Rigg, and G.N. Kechkhuashvili, in which the character of the musical experience, and especially of the process of musical–auditive activity in general, were determined by externally formed "meaningful" framing, in our experiments the process of perception of emotional message of the music by the listeners was determined by "preset" framing, fixed and expressed in their own movement only in the process of realization of activity. The composition of emotional–meaningful structures of the image was carried out not with the help of externally induced cognitive programs, but on an emotional basis; In the conditions of dynamic transformation of musical–aural frames connected with the perception of musical information. More static "visual" frames connected with the perception of different variations of external expression of a performer have a blocking effect on the process of development of identification, prognosis, decoding and audio–perceptual acts and operations.

2) The visual (additional) information was more significant for the majority of the listeners than the aural (main) information. The experimental results supported the hypothesis of the exceptional role of visual factors in perception of telecast music.

3) The results have social–psychological and psychological–pedagogical values. As regards the material on musical activity, the effects of individual categories (including age) of a performer and of the character of emotionality on the productivity of social–perceptual activity of a listener were shown.

4) The developed technique can be used as a test for the assessment of the level of musical culture of a listener. The results of the experiments testify to the necessity of a more specific attention to the formation of aural skills for comparative assessment of the character and message of interpretation in education of future performers, teachers, and members of boards of adjudicators at musical competitions.
Perceiving intended emotion in concert-standard performances of Chopin's Prelude No. 4 in e-minor

J. A. Sloboda, A. C. Lehmann, & R. Parnscutt
Keele University & Florida State University

Introduction

Music performances create a range of intellectual, aesthetic and emotional responses in listeners. These effects are, to a greater or lesser extent, intended by performers and composers. In investigating musical expression, we can therefore focus on different aspects of the communicative process that extends from the composer to the listener. We can try to explain a listener's response to music by (a) considering characteristics of the listeners themselves such as their personality, training background, and preferences, (b) trying to predict listeners' responses based on observable indicators of the structural properties of the performed music, or (c) relating a listener's response to observed characteristics of a particular performance. This study is a first attempt to investigate all three of those aspects and their relationships.

With regard to structural properties of music, listeners may parse the performance so as to assign metre, phrase structure and other forms of segmentation. Other effects of performances on listeners have not so much to do with how the music is divided up into units, but more about specific qualities ("characters") of the music that a listener detects or feels. Thus, at any particular moment, a piece of music may be experienced as more or less tense (Krumhansl, 1996) or more or less pleasant (Madsen, Byrnes, Capperalla-Sheldan, & Brittin, 1993). In these latter two studies, listeners made continuous adjustments to a sliding scale or dial to indicate changes in intensity of the quality being examined over time. Both studies revealed a satisfactory degree of inter-rater agreement concerning the measure-to-measure changes in characteristics such as perceived tension (also Gregory, 1995). In addition, Krumhansl demonstrated how tension estimates could be related to indicators of the musical structure (also Tyler, 1996).

The work on continuous ratings during music listening has, to date, not compared different performances of the same piece. It is not possible, therefore, to separate effects of musical structure from effects of the particular performance, and to assess effects that are listener-related. A number of investigators, notably Repp (e.g., 1990) has examined the effects of different professionally recorded performances of the same piece on listeners. He has shown that objective characteristics of the performances, such as differences in absolute and relative timing of notes, reliably affect judged characteristics of performances. In his studies, listener judgments were collected by means of post-performance bipolar rating scales. These scales assessed what is sometimes called "performance character" and include such dimensions as cold-warm, smooth-rough, powerful-weak, soft-hard. Repp found that changes in the way a piece is performed can very substantially affect its perceived character.

It is well established that listeners within a culture tend to agree with one another about the emotional character of a piece of music, as evidenced by their choosing closely similar adjectives from a check-list to describe the music. Until recently, rather little scientific attention has been paid to the task of tracking moment-to-moment emotional response over the course of a piece. The present study was motivated in part by the recent demonstrations (Sloboda, 1991, 1992, Waterman, 1996) that it is possible to link peaks of emotional response to quite specific moments in a musical experience (see also Tyler, 1996). For instance, Sloboda (1991) asked music lovers to recall pieces of music which reliably provoked the physical concomitants of strong emotion such as crying and shivers down the spine (often called "thrills" or "chills," see also Panksepp, 1995). Many respondents were able to identify short sequences within a larger work which were the reliable focus of their response. These sequences tended to share music-structural characteristics, such as appogiaturas and suspensions, and harmonic change. However, findings were normally
related to a specific performance or recording of a piece, making it impossible to systematically probe differential effects of piece and performance.

It seems very likely, from personal experience and anecdotal report, that different performances of the same piece can have very different levels of emotional impact for reasons that may or may not be connected with characteristics of the individual listener. The present study aims to discover how score-based structural indicators, observed performance parameters, and listener characteristics influence the emotional response to music. Composer and performer make decisions which heighten or diminish effects that the composition can create in listeners. Also, we are interested in assessing individual differences across different groups of listeners with regard to their on-line responses.

Method

Gathering performances. Participants were 10 volunteer members of the solo performance department at the Sibelius Academy of Music, Helsinki. They were either staff or advanced graduate students of the department, and were all fluent English speakers. Also, all had professional concert playing experience. The music to be played was F. Chopin's Prelude No. 4 in E minor (Op. 28, No. 4; see Schachter, 1994, for a description). This particular piece has been used previously in music research (e.g., Clarke, 1995), likely because it contains high levels of emotion, is not technically difficult, and is relatively short. We expected that professional pianists would have no difficulty preparing an error-free performance with limited preparation time. All participants worked from a score with all editorial fingering marks removed, but the composer's phrasing and expression retained. After a one month preparation period, the pianists were asked to give two concert-standard performances of the test piece, to be as similar in interpretation as possible. If the participants were unhappy with either performance, a third performance was invited, and the participants could then choose the two that they were most happy with. MIDI recordings were made and used in a subsequent listening experiment.

Subjects. Participants were 28 Finnish musicians (all fluent in English) belonging to one of three groups, defined on the basis of their pianistic expertise. Group 1 consisted of 13 non-pianist musicians. These were primarily members of the Sibelius Academy who, although they might have some amateur pianistic ability, had their primary musical expertise in another area. Group 2 consisted of 7 professional jurors, who regularly judged piano performances in competition settings. Group 3 were eight of the ten pianists who had provided the performances (see above).

Materials and apparatus. The stimulus materials were the 20 performances of the test piece (see above). Participants of the listening experiment heard resynthesized sound generated from the stored MIDI files, rather than acoustic recordings. The playback equipment used recreated the original sound quite accurately and was the same for all listeners. However, each participant adjusted the sound level for comfort, but the level then remained constant for each participant. Timing and order of presentation of performances was controlled by computer program written by one of the authors (AL). The program also provided the control structure for the on-line presentation of instructions to participants and the collection of their response data.

Design. Piloting had shown that 12 performances of the test piece was the maximum than could comfortably be heard within a one-hour testing session. Accordingly, the 20 performances were divided into two sets of 10 on the basis of speed of performance. Set 1 contained the performance of the five faster pianists (mean of duration of 94.4 secs, SD = 7.4), while Set 2 contained the remainder (mean duration 108.9 secs, SD = 8.8). The tempo difference between sets was statistically reliable (p < .05). Each participant heard one complete set. The rationale for dividing the performances in this way, rather than randomly presenting 10 was to avoid gross differences in expressive quality brought about solely due to widely differing speeds. Within each expertise group, as close as possible to equal numbers of participants were presented with each set of performances. Each participant heard the 10 performances of the relevant set in a different random order.

Collecting response data continuously. Each participant was tested individually in a quiet computer music studio in the Sibelius Academy, Helsinki. Instructions were presented on the computer.
console, and reinforced by the verbal commentary of the experimenter. Subjects were asked to indicate on a horizontal scale how much emotionality was present in the performance from moment to moment. Emotionality was explained as the capacity of the performance at that moment to suggest, communicate, or evoke musically relevant emotion. The main instructions were also translated in Finnish as it was important to confirm that these key instructions were well understood. The subjects moved a mouse cursor in a response box, which was calibrated for participants by 11 equally spaced markers at the top, labeled 0 to 10 from left to right. The far left (0) point was labeled “minimal possible emotionality”, the far right edge (10) "maximal possible emotionality". The software was designed to collect x-axis positional data of the cursor at every 6th note of a performance. The numbers returned by the software ranged from 0 (left edge) to 300 (right edge). The sampling points for an individual performance were always the same across performances and very similar to the sampling points in other performances. Some 100 samples were taken for the 600 notes contained in the piece.

Post-performance ratings of the interpretations. After each performance, the participants rated onscreen the overall performance on 10 bipolar scales: expressive-inexpressive, superficial-deep, powerful-weak, spontaneous-deliberate, coherent-incoherent, sloppy-precise, restrained-excessive, rigid-flexible, interesting-boring, and effortful-effortless. The first eight of these were selected from the dimensions used by Repp (1992) in a study of listener-ratings of performances of a Beethoven minuet. The last two dimensions were added to address further characteristics of a performance that are likely to affect its emotional appeal.

Results

The data analysis addressed three different questions. First, what relation exists between responses that are made on-line while listening to a piece of music and summative ratings made on post-performance bipolar ratings scales? Second, do different groups of expert listeners show similar responses when listening to the music? And third, can we predict continuous emotionality ratings using objective indicators of musical structure and observed performance parameters of the different interpretations?

Relation between post-performance ratings and continuous measurements of emotionality

The first section addresses the possible relationship between the continuous responses listeners gave while they were listening and the ratings they gave retrospectively after listening to the entire piece.

Continuous Emotionality Measurements. When averaging across all 100 sampling points and across all 20 renditions the mean emotionality response was 177.1 with a standard deviation of 44.6. Figure 1 shows the general outline of the response curve over the 100 sampling points. The curve follows a path that could be expected based on the musical structure with the highest point toward the end of the streto part (measure 18). The initial increase forms a ramp until around measure 4 where the first plateau occurs. After that, there is a gradual increase toward the apex, after which the curve descends to roughly the same level that it had at the first plateau. After a visual inspection of the overall curve, it was decided to divide the piece into four shorter segments that would preserve the general outline of the response curve and be related to the musical structure of the piece. The beat demarcations for the segments (Segment 1 [beginning - m8, b2], Segment 2 [m8, b2 - m15, b4], Segment 3 [m16, b1 - m20, b2], and Segment 4 [m20, b3 - end]) are approximate since neither the sequence of keystrokes nor the exact number of notes in every piece are identical. A repeated measures ANOVA showed that segment means were significantly different, multivariate F(3,265) = 147.80, p < .001. Post-hoc tests between adjacent segments were also significant, showing that the different segments in fact captured different levels of perceived emotionality.

Post-Performance Ratings. The 10 ratings scales were subjected to a principal component factor analysis. This was done to arrive at a simpler description of the listener’s impression, which could later be related to the continuous responses. This factor analysis revealed an underlying structure with three factors with eigenvalues greater than one. Together, Factor 1 (IMPACT), Factor 2
(PREPAREDNESS), and Factor 3 (EXAGGERATION) explained 73.9% of the variance in the ratings. Further analysis were done on the saved factor scores.

**Correlations between segments and factor scores.** The correlations between segments and factor (see Table 1) scores are mostly reliable. However, the correlations for the factor IMPACT are generally higher than those for the other. Higher values on this factor indicated more positive attributes regarding expressiveness, perceived power, spontaneity, coherence, and interest. A subsequent regression analysis for the overall emotionality rating across all sampling points showed that when allowing all three factor to enter, the unadjusted multiple R is .51, F(3,266) = 31.79, p < .001.

**Table 1. Correlations between individual segments and factor scores.**

<table>
<thead>
<tr>
<th>Factors / Segments</th>
<th>Segments 1</th>
<th>Segments 2</th>
<th>Segments 3</th>
<th>Segments 4</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impact</td>
<td>.23**</td>
<td>.43**</td>
<td>.47**</td>
<td>.48**</td>
<td>.45**</td>
</tr>
<tr>
<td>Preparedness</td>
<td>.09</td>
<td>.09</td>
<td>.17*</td>
<td>.17*</td>
<td>.14+</td>
</tr>
</tbody>
</table>

Note. n = 269; +p < .05, *p < .01, **p < .001

Thus, although the post-performance ratings showed a strong relation to at least one of the factors, they only explained about 25% of the variability in emotionality ratings. The next question concerned the ratings of the different subject groups.

**Influence of subject group on ratings and continuous responses.**

The three groups of participants differed from each other with regard to the expertise they had with piano performances. It was assumed that the jurors were the most expert group, followed by the pianist, followed by the musicians who were non-pianists. The first analysis was intended to assess possible differences among the three groups with respect to their on-line ratings of emotionality. A MANOVA with a repeated measure on the first factor (Segments 1 through 4) and group as a between subjects factor showed a significant main effect of segment (F[3,254] = 149.5, p < .001), a significant effect of group (F[2,256] = 8.5, p < .001), and a reliable interaction segment x group (F[6,510] = 7.21, p < .001). Although the interaction clearly attests to the different trajectories of rated emotionality between the groups, the main effect of group may have been due to varying ranges of the scale being used by the participants. Therefore, each segment score was standardized using the overall score and standard deviation of a given subject for a given performance. The analysis was repeated for the standardized segment scores. Again, both main effects and the interaction were significant (see Figure 2). Subsequent post-hoc tests (corrected for multiple comparisons) showed significant group differences for the first and the last segment, but not for the two middle segments. Compared to their own average ratings, pianists started out reliably lower than the other two groups, but finished reliably higher than the other two groups. Jurors and non-pianists did not differ significantly in their ratings after standardization. To assess possible differences on the post-performance ratings, a MANOVA was conducted with the three factor scores as dependent variables and group as a between subjects factor. The multivariate effect was not significant (power .65) and the univariate tests for each of the three factor scores were also not significant. Thus, group membership did not seem to influence the mean post-performance ratings reliably.

**Predicting emotionality ratings from structure and observed performance.**

In an ongoing effort to investigate the interplay of structure and performance characteristics we are trying to explain changes in the continuous emotionality ratings. The last analysis is therefore a first attempt to explain changes in emotional intensity throughout a given performance. At this point, there is no theoretical description which could capture the temporal summation of changes in the musical structure. However, we can describe individual points in the music and their structural
components. In addition, we can extract indicators from the observed performance such as timing and loudness which may influence the listener's experience.

For the following analysis two pieces were chosen based on their standard deviations across all participants. The piece with the smallest (example MO2; SD = 39.27) and the piece with the largest standard deviation (example KH2; SD = 59.97) represent interpretations where listeners' showed rather homogeneous and heterogeneous ratings, respectively. For those two pieces, observed loudness (velocities) and timing between quarternote beats in the bass were extracted from the MIDI data. The length of the inter-onset interval between bass notes are essentially responsible for perceived expressive changes in tempo. In addition, a description of melodic and harmonic accents was developed by one of the authors (RP; see also Parnscutt, 1997). A regression analysis was conducted in two logical stages. While the structural indicators were entered in the first stage, observed performance variables could account for additional variance in the second stage. For the piece with the smallest standard deviation, only the direction of the melodic intervals in halfsteps was a reliable predictor and explained 18% of the variance, F(1,47) = 11.74, p < .001. No other variable entered the equation. The results for the piece with the largest standard deviation are somewhat different. Here, the direction of the melody accounted for 13.7% of the variance. Bass note timing added another 23.7% of variance, adding up to a total of 36%, F(2,44) = 14.28, p < .001. The observed velocities failed to enter the equation (F = 3.99, p = .052). Thus, the continuous emotionality ratings in the more homogeneously perceived piece were significantly influenced only by structural properties, while the more divergently perceived interpretation showed some contribution of unique performance aspects.

Figure 1. Mean response curve and standard deviations for continuous measurements of emotionality across all renditions and subjects.

Figure 2. Separate group means and sample mean for the four segments.

Discussion
This study combined the traditional post-performance assessment of a musical listening experience with the on-line measurement of listeners' emotionality ratings. Its purpose was to shed light on the influence that listener characteristics, objective musical structure, and performers' unique interpretations exert on the perception of emotional expression in piano performances.

Post-performance ratings were factor analyzed, and one factor (IMPACT) correlated more strongly with the concurrent measurement of emotionality than the other factors. Incidentally, this factor contained some rating scales with adjectives similar to the one offered to the subjects for the continuous measurement. In general, these results are consistent with those found in other studies on continuous response measurements in music listening (Gregory, 1995). Furthermore, it was
shown that listener characteristics such as expertise with the performed music and the instrument on which it is performed may influence the moment to moment experience. When the piece under investigation was divided into four segments, the group of pianists was found to differ reliably from the other groups. However, no significant group differences were found for the middle segments which suggests that the beginning and end of a piece might contain crucial information that subjects interpret differently based on their prior experience. That no significant effects of expertise were found with regard to the factor scores for the post-performance ratings awaits further explanation, especially since expertise effects existed for some of the rating scales when analyzed individually. Finally, although our results are only tentative at this point, it appears as though the structural aspects of the music as well as the observed indicators of performance work together to create the expressiveness that our listeners perceived. Most surprisingly, none of the observed performance variables (inter onset intervals between bass notes and velocities between sampling points) was a reliable predictor in the performance with strong listener agreement. Yet, they contributed to the explained variance in the performance with larger variability in continuous emotionality ratings. Further analyses might show that structural properties take precedence over unique features of the interpretation in cases where listener agreement is strong.

In essence, our results are in agreement with Clarke (1995, p. 50) who succinctly points out that the listener is likely to perceive a number of small cues that by themselves may not be very salient, but that add up to an impression that is more than the sum of its parts. It may be that the musical structure is initially more salient, while nuances of the performance require time to be noticed, and they may even require prerequisite knowledge and experience to be fully recognized.

References


Teenagers’ experiences in typical “aesthetic areas”

Leif Finnäs
Pedagogiska institutionen
Österbottens högskola, Vasa, Finland

Background

In traditional music psychology, music experiences have typically been studied in experiments using brief music excerpts and focusing on how one or a few predetermined aspects of listeners’ reactions are affected by the researcher’s manipulations of some aspect of the tonal material. Also when intact compositions have been used in experiments, or when in non-experimental research subjects have been asked regarding their reactions to “real” music (including types of music, and music in general), researchers’ interest has mostly concerned only a few specified variables, such as the subjects’ preferences, tendency to visit concerts, and the like. Such approaches may have notable advantages as it comes to conceptual clarity and methodological stringency, but especially in recent years they have also been criticized for their relative inability to describe and explain persons’ many-faceted reactions when they normally confront music. Such critical considerations can be found among the starting points for some studies aimed at investigating real-life music experiences in their totality. In these studies, subjects have been asked to describe in detail naturally occurring musical events which have made some special impact on them. In a context relating to childrens musical development, for example, Sloboda had adults write down memories involving music from the first ten years of their life. Panzarella, and more recently Gabrielsson and his collaborators, have collected persons’ descriptions of especially intense experiences, the former one speaking of musical “peak experiences”, the latter of “strong music experiences”, understood to belong to the clearly most intense music experiences ever had by the individual (Sloboda 1990; Panzarella 1980; Gabrielsson 1989; Gabrielsson & Lindström 1993).

The current study

General purpose of the project

The ultimate motive behind this project was an interest of the kind suggested above, to describe persons’ real-life music experiences in a comprehensive way. On the whole, the project is characterized by three distinctive features. Firstly, the subjects are young people in their middle teens, an age category not separately singled out for study in earlier research of this kind. Secondly, the main focus is on “typical strong experiences” which may occur in the young persons’ daily life, and which need not necessarily be of the “peak” type, or uniquely intense. Thirdly, the original research idea concerning music experiences was completed with
the idea of simultaneously applying this approach also to the teenagers’ reactions in other areas which are generally considered as typical “generators” of aesthetic experiences, i.e. to reactions emanating from contacts with art (including film and theatre, dancing etc) and literature, and with nature.

Due to the explorative and “qualitative” character of this project, no hypotheses or narrowly specified research questions are formulated in advance. Of evident interest will be the assessing of the relative importance of different experience areas (nature, literature, music etc.) for these pupils, to describe different aspects of the experiences (perhaps applying dimensions similar to those used by Panzarella and Gabrielsson) and to analyze the conditions contributing to the impact of the experiences.

Subjects, procedure

The subjects are ninth-graders (age: 15-16) in the Finnish obligatory school. The collection of data has been going on since 1993, and will probably continue until the academic year 1998-99, when data for about 500 pupils have been gathered. Data are collected in class during one lesson in Swedish (these pupils’ mother tongue). Subjects’ main task is to write a short essay about (at least) one “strong” experience they have had in some of the above-mentioned areas, thinking mainly of rather recent experiences, even within the last month or so. They are told to describe the experiences as such and their situational context, to speculate about the reasons for the “strongness” of the experiences, and to assess their intensity on a simple rating scale. Through lengthy instructions and exemplifications, it is tried to make clear to the pupils what kinds of experiences they are expected to report and not to report (not, e.g., strong experiences connected with tragic incidents, with love affairs and friendship relations, with one’s own or others’ successes or misfortunes in school and sport activities, etc.).

Some preliminary results

For two aspects of the experiences which can be readily assessed and which are of interest from the point of view of music psychology, some preliminary results are presented in the tables. Regarding the distribution of pupils’ experiences in different areas, some tentative conclusions regarding the main trends could be made from the data in table 1, based on a sample of intact classes, chosen to get some minimal representation of subgroups considering both sex and degree of urbanization. The category “other” in the table comprises a few experiences concerning e.g. sculpture and painting, but mostly represents irrelevant experiences (personal failures, achievements of the Finnish ice hockey team, etc.). On the whole, music and nature experiences seem to be the most frequent types of relevant experiences, with music experiences being clearly most frequent among the “art-related” experiences (note that mentioning of music only as a background factor not further commented, did not here qualify an experience as “musical”). Regarding subgroups, one notes the boys’ greater tendency not to report relevant experiences or any experiences at all, and the peculiar absence of reported nature experience among urban boys.
Table 1 Frequency of ninth graders' reported experiences from different areas (with n=124 pupils, i.e. generally only one experience for each pupil).

<table>
<thead>
<tr>
<th>Town</th>
<th>Girls</th>
<th>Boys</th>
<th>( \Sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Music</td>
<td>11</td>
<td>12</td>
<td>23</td>
</tr>
<tr>
<td>Literature</td>
<td>6</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>Film/thea</td>
<td>5</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Nature</td>
<td>8</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>( \Sigma ) 33</td>
<td>( \Sigma ) 33</td>
<td>( \Sigma ) 66</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Countryside</th>
<th>Girls</th>
<th>Boys</th>
<th>( \Sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Music</td>
<td>11</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>Literature</td>
<td>6</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Film/thea</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Nature</td>
<td>10</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>( \Sigma ) 33</td>
<td>( \Sigma ) 27</td>
<td>( \Sigma ) 60</td>
</tr>
</tbody>
</table>

Table 1 Frequency of ninth graders' reported experiences from different areas (with n=124 pupils, i.e. generally only one experience for each pupil).

Table 2 specifically concerns music experiences, indicating the medium (in broadest sense) through which the music was “produced” (most of the 52 experiences belong to the just-mentioned sample). Half of all experiences are aural ones, with half of the remaining experiences resulting from the pupil’s own performing of music.

Concluding comments

This exposition has been aimed at presenting an ongoing research project and a few preliminary findings, and definitive generalizations should of course not be made until comprehensive analyses of complete data have been carried out. The above data, however, seem sufficient to indicate music as a very important “experience area” for this age category, also when compared to the other experience areas chosen for study. The relative importance
of the different experience areas certainly has to do with external factors such as the practical accessibility and the general “base-rate occurrence” of events in the respective areas, as well as with social factors and with factors concerning how events in the different areas intrinsically exert their effect on the young person. Closer qualitative analyses of pupils’ responses may shed some light especially on the contribution of the latter kind of factors. Such reflections pertain also to the “medium-type” character of the experiences, where it could perhaps be especially interesting to note that the frequency of reported audio-visually mediated musical experiences appears to be very low, although for children of this TV- and rock video era the base-rate occurrence of audio-visually produced “youth music” is very high (a short questionnaire confirmed that most pupils watch music videos every day). This observation seems to parallel certain earlier findings by the author and others, suggesting that using the video/TV-medium for exposing young persons to some music may often not have any enhancing effects on their music experience (Finnäs 1992).

Apart from their theoretical interest, the results from a project like this could of course have obvious practical implications, especially perhaps in educational contexts. For example, knowledge about the distribution of strong experiences over different experience areas may imply conclusions about the importance of treating the respective areas in culture and art education, and suggest ways of utilizing more easily aroused kinds of experiences (perhaps regarding music or nature) to “facilitate” experiences in less frequently experience-generating areas (perhaps regarding painting, sculpture, architecture). Knowledge about the efficiency of different media to enhance teenagers’ music experiences clearly has implications for the ways in which a music teacher confronts his or her pupils with some type of music.

References


MUSIC IN THE CAR
Carin Öblad
Department of Musicology, University of Göteborg

Introduction
Music and cars represent strong interests in society. It is common knowledge that music is often played in the car, but research literature relating to this phenomenon is limited. The aim of the project "Music in the car" is to focus on how an individual in her everyday life uses music for different purposes with special regard to the situation driving a car. The empirical material consists of a quantitative mapping out and qualitative interviews.

Pilot study
In the pilot study The Automobile as Concert Hall (Svenson, 1992), seven individuals having different frames of references were interviewed. The main results implicated that the music a person wants to play in the car should be rhythmic, vocal and familiar. The preferred genres were popular hit songs and varieties of rock music. The use of the music was either a conscious or an unconscious choice: to create an atmosphere, to stay awake, to have company, to feel good, to get physical and mental energy, to reduce unwanted noise, to achieve individual body rhythm and to reach an alternative reality or to escape reality.

Some individuals were familiar with their own reactions and chose their music in order to use it in different situations. One woman, who almost solely listened to classical music at home, played disco music when she cleaned the house or in the morning when driving to work. She did not, however, dare to play "I feel love" by Donna Summer any longer since she had found out that, when playing this song, she simply drove too fast. She also knew which ballads suited her best when driving home after work, feeling tired.

A young woman expressed that she preferred to take her car to and from work in spite of traffic congestion, rather than taking the subway.
"You sit there calmly, listening to music or smoking a cigarette. It's more relaxing than travelling by public transport...you can think and you are left in peace for a while. As a matter of fact, I think this is quite nice".

Licentiate thesis
In the licentiate thesis Music in the car. Actions and preferences – On the Car as a Concert Hall (Oblad, 1996), the above quotation formed the incentive for three themes laying the basis of an extensive questionnaire:
To be left in peace and to be left alone
The car as a life sphere and as a room of your own
Using the music to manipulate yourself.
A comparison of how music was used in the respondent's home and in her car was central in order to find out what the car can offer as environment for a musical experience that the home cannot.

Questionnaire and On-going interviews
Following up the pilot study and the questionnaire, interviews are now carried out with individuals selected among those who answered the questionnaire. Focus is on How does a driver use music in the car? and the concept concerns the experiences of the chosen music in the car. Even at this early stage of the analysis, a pattern is obvious, that one of the strongest reasons to play music in the car is a wish to get access to memories. The music evidently opens the memory of the situation with the attached feeling and the mood, when the music was played. The medium most frequently used is tapes produced at home and consisting of the "goodies", the best parts of CD records. These tapes are played over and over again, especially at long trips.

In this paper I will present results from the questionnaire and some from the interviews (the analysis of these are being done at present). However, the incentive for the questionnaire and the discussion of the results are strongly related to the pilot study (mentioned above) and the licentiate thesis, why these – other literature related to the project is very limited – will play a role as a basis for the discussion to follow in this paper.
Method
The questionnaire was answered by 388 Swedish individuals representing different ages (25 and 53 years), gender, and education. The formulation of the questions was done in collaboration with Statistics Sweden, which also carried out the statistical processing of the answers. This data set has been analysed and interpreted by me. The purpose with the questionnaire was really twofold: To survey actions and preferences and to provide the basis for the selection of informants to be used in the upcoming interviews. Eight questions were free and concerned what kind of music the person liked and disliked and the reasons for this. The other 57 questions had fixed, alternative answers. The questions were divided into six groups, related to people's everyday life: Living conditions; the Car; Music and Media; Leisure; Preferences and Activities; Existentialism and Emotions.

The ongoing interviews are carried out following an interview-guide constructed by me, which provides the study a frame as well as flexibility. The questions are variations on the theme Use and function of music with special regard to the environment of the car. The interviewees are selected amongst the respondents from the questionnaire by strategic choices: to balance the different answers to the questions 61 and 59 (presented later in the paper) in order to avoid reasoning in a circle. Considerations are also taken to strata and residence in Sweden.

Results and discussion

Results of the questionnaire
"To be left in peace" can be interpreted in a number of ways: As an escape from a troublesome existence (e.g. problems at work and at home) or as a conscious change of reality/environment. Especially the younger generation differentiates between "to be by yourself" (to voluntarily spend time alone, for your own sake) and "to be alone" (to feel alone and/or involuntarily be without company).

I especially interpret "to be left in peace" as an expression of wish for an abstract integrity, which can be separated from the concrete isolation and interpreted as having an existential value. The interior of the car may be experienced as a living sphere, a private room where the person in her daily life creates shorter or longer sequences of existential value. In this context, music has a function with obvious effects.

I have expressed it as follows: She increases her sense of presence in life which is a recovery or an intensification of herself as an individual. These experiences can be of different sorts and intensity, e.g. ecstasy or peacefulness, and may be viewed as temporary breathing spaces hidden in a trivial context like automobile driving and may function as existential edification.

A central question in the questionnaire was:
61) If you want to be left in peace, do you think the car is a good or a bad place?
- Good
- Bad
- Don't know/Have never thought about it

51% of the younger persons and 45% of the older answered "good" while the equivalent percentage for "don't know/have never thought about it" was 32% and 39% respectively. These answers should be viewed in relation to the question whether or not there is a room at home where they can be left alone and do whatever they want. Approximately 85% answered that they have such a room. These results indicate that the car has an autonomous value in relation to the home, a value that the home cannot offer.

Younger as well as older people are of the opinion that they can play whatever music they want at home and as loudly as they please. This also indicates that the car is not an alternative to something you can't do at home, e.g. listen to music at a high sound-level.

Another central question was:
59) Where do you get your most intense music experience if you only compare the car and your home?
- The home
- The car
- No difference
- Don't know/Have no opinion
The data is evenly distributed across all three alternatives. The respondents who chose the car as the environment for the strongest musical experience also to a great extent consider the car as a good place for being left in peace. They have stated the following effects of the music: Stress, irritation and tranquillity, relaxation and changes in the mood. They are quite aware of how they react to different sorts of music at home and in the car.

Every experience is unique since each individual has, for example, her auditive qualifications and her private memory connected to a certain piece of music. Man can function as a chameleon by listening to the same piece of music but in different ways and in this way use the music. She can use different modes of listening in different situations. This complexity in a musical experience is something which only the individual listener knows about and, hence, only can understand herself.

One more question can illustrate how music has a function in this context. The question was: 57) What do you do to stay awake and keep alert when you are tired but have to drive another 50 kilometers? Name the alternatives you normally choose.

☐ Smoke
☐ Play music
☐ Eat sweets or other things
☐ Open the windows
☐ Stop the car, walk about and continue driving
☐ Other ...........................................................

Two alternatives got the same, high answering frequency: "Stop the car, walk about and continue driving" and "Play music". In this situation it is shown that, by choosing "Play music", the music has a function - to counteract fatigue.

The results from the questionnaire show that the car is more than a utility. It also responds to the need to be left in peace as well as fulfill desires e.g. experiences in connection with listening to music especially.

Some of the explanation to the failure of car pools can probably be found here. If you want to be left in peace in traffic congestion in order to have time on your own in a natural way, maybe to be nostalgic and play the music which suits your frame of mind, then another individual will be perceived as a disturbing factor. To give general advice as to which music, which composer or which artist is suitable when driving a car is a doubtful measure. The musical experience is too unique, especially when it affects the person strongly.

Results from on-going interviews

The results from the on-going interviews have thrown a more revealing light on the answers from the questionnaire. For example, one reason for wanting to drive alone in the car is a desire to sing: a number of informants categorizes themselves as "car singers". They also declare themselves as being poor singers, thus preferring no one to listen while they sing the familiar songs from their favorite tapes.

Depending on the kind of activity at home or in the car, an individual decides either to play or not to play music: domestic work, getting access to feelings, to get relaxed are examples when music is wanted. Studying or solving theoretical problems are situations when music is not wanted. I see this as an expression of "use of function": the activity demands that a certain function of the music is exerted on the individual.

Certain strong experiences of music are attached to an event, but at this stage of analysis, it seems almost impossible to separate the event from the music: does the informant want to recollect the music or the event or both?

A consequence of the unique and private character of the musical experience is that one must be careful not to generalize as to the effect of different kinds of music. The difference between generations was obvious in the questionnaire when it comes to musical preferences: The older people are much more negative to "hard rock" and positive to "ballroom music" and "classical music". Many older people feel stress when they hear rock music and some feel harmonious when listening to Mozart. The younger men, on the other hand, get stimulated by rock music and even mention that the music by Mozart is foolish which makes them irritated. These different experiences of various types of music either add to the distraction or increase the concentration.
There is a clear tendency amongst the interformants concerning classical music which differs from the above concerning respondents. Almost unanimously, the informants have declared that listening to classical music makes them feel calm, at peace. The factual knowledge of composers or work titles is vague and almost limited to just "classical music". The composers being mentioned are mostly Vivaldi and Mozart. This music is bought as compilation CD with classical hits.

**Use & Function**

With the concept Use & Function, it was argued by Merriam (1964, p 210) that "'Use' then, refers to the situation in which music is employed in human action; 'function' concerns the reasons for its employment and particularly the broader purposes which it serves". The distinction was that the layman was only conscious of 'use': folk evaluation, whilst 'function' could only be understood and described by the researcher, the outside observer: analytical evaluation. Amongst others, Nettle (1983) has in the discourse pointed out the difficulties to separate use from function. Nettle also criticized the research ethnocentric perspective by Merriam concerning function. The ethnomusical concept discusses the collective and this generalistic holding has been criticized by Jørgensen (1988) stating that the functions of music are unlimited and strongly connected and related to the individual. In *Strong experiences of music* (Gabrielsson & Lindström, 1991) the qualitative part consisted of the verbalised descriptions of individual experiences of music. In a later article, the authors put forward the question if strong experiences of music can have therapeutic implications (Gabrielsson & Lindström, 1995). As I see it, the music has in this context a function known to the listener, that is, the individual who had this strong experience of music. This was demonstrated by writing the description.

I would try to apply the concept *use of function* since the individuals in my studies seem to be conscious about the effects of music. When the interviewees are asked about the process of choosing music in a specific situation, the following answer is frequently given: I never reflect upon this, but I never do wrong. How does this process of choice work? Evidently it is a choice including knowledge of the effects of music, that is, the function of music. Is it possible that the concept "tacit knowledge" is applicable here? And how can it be made visible?

**Future research**

Some re-interviews will be carried out. It is evident that the interview itself with its questions concerning feelings and effects, has influenced the informants. Comments such as "You make me think of my reactions", "Now I will go home and play the music we have talked about" are common. There is a possibility of increased, individual reflection upon the matters as a consequence of the interview situation. If so, it would be of interest to meet the informant again. I have instructed but one interviewee to reflect upon himself while he was aware of what feelings were connected to certain music. The intentions are otherwise to contact the informants unexpectedly for a re-interview.

Field experiments are planned with car trips on country roads and in urban traffic with one or more of the interviewees. The process will be recorded on video and different sound types (dialogues, comments). The intentions are to manipulate the car driver with music. The driver is then supposed to comment the effects in addition to that which can be registered with the various equipment. A cooperation has been established with VTI (the Swedish Road and Transport Research Institute) which has constructed an experimental car enabling measurements of different kinds. This is an alternative to the car simulator; thus using the experimental car there can be a reasonable chance of ecological validity in the aspect of letting the driver drive alone in the car if yet not drive a car of her own.

**Acknowledgements**

I want to thank my supervisors Stig-Magnus Thorsén and Alf Gabrielsson for their patience and imagination. The investigation is supported by the University of Göteborg and KFB (The Swedish Transport and Communications Research Board) and through scholarship from The Royal Swedish Academy of Music.
References


Introduction and theoretical background

Starting point is the assumption of a fundamental analogy between the development of language metaphors and the process of music reception. This analogy between the language and the cognitive process, generally spoken, calls for an integration of qualitative experiments into the investigation of music cognition (Haskell, 1986).

Music sounds like emotions feel. (Langer, 1942) In an analogue way metaphors demonstrate how perception operates.

Johnson (1987) reflects „the bodily basis of meaning, imagination, and reason“. Lakoff (1990) uses the metaphor in the sense of a systematic mapping between a source and a target area. Whorf (1956) considers the nearly magic potential of language to enlighten and to get conscious of certain dark psychic impressions. Only a little further step leads to the music experience which may communicate a large stream of general feelings, thoughts, and imaginations beyond reality.

The traditional meaning of metaphor has to be revised. Its definitions reach from an objectivist’s to an experientialist’s or contextualist’s understanding (see Lakoff, 1990 and Gillespie, 1992). Metaphors no longer may be considered as pure rhetoric ornaments. In place of that metaphors may communicate a deeper understanding of music experience. Moreover, they create certain meanings of music on the way they arise. So the metaphoric process in general might be valid for essential insights into the nature of music experience.

One main result taken from earlier investigations is a hierarchy of verbal categories on four levels and a typical bottom up process from bodily sensations via general image schemes and language metaphors up to musical thinking (Kleinen, 1994, 1996).

Seiler and Claar (1993) take it for granted that the beginnings of a verbal term lie before the term. First only isolated images exist fixed at concrete subjects, specific situations and events. By abstraction, generalisation, also by the building of complexes, co-ordination and reflection the isolated images change to verbal terms in its real sense (p. 108). In a constructivistic conception verbal terms are not taken over as ready made and from outside, not in an unidirectional sight, but as a construction undertaken by the subject itself. „What a subject may receive, how it receives and classifies, how it understands the words of an interacting partner, depends from the activity of its structures.“ (p. 117) „Without any doubt the terms mirror experienced reality and at the same time they reflect sociocultural aspects“ (p. 117). There is only a little step to the role of metaphors for the integration of single events into the whole life experience.

Already Piaget (see Piaget & Inhelder, 1966) observed that the senso-motoric intelligence is fundamental for the construction of reality by children and youth people and also as presupposition for cognitive and affective reactions. The development of perception and the growing of the semiotic and symbolic functions all base upon senso-motoric experiences of the body.
The leading questions for the following experiment are: What does the „body in the mind“ mean with regard to music experiences? How do metaphors grow up in the musical context. How powerful are metaphors to explain music experience? Which repertoire of metaphors is essential, which general image schemes and which bodily experiences may be originated by the music? What are the structure and the contingencies of metaphors belonging to the music? Which are the typical down-up-processes? What are the benefits of language analysis in respect to the music reception processes?

Procedure

For the investigation I performed a qualitative experiment. The procedure was to ask for verbal reports under the immediate impression of the music. A series of music examples which may be characterised by different familiarity and strangeness (identification and distance). The sample consisted of two pieces of Western music (Beatles: All my Loving, 1964; Wolfgang Amadé Mozart, Adagio from the piano sonata in B-major, KV 570, 1789) and two of Asian origin (a piece of Chinese folklore from the ethnic minority of the Dan, played on a bamboo-flute and accompanied by traditional Chinese instruments; an avant-garde orchestral composition by Tan Dun „Ad parrasum“, taken from the cycle „Death and fire - dialogue with Paul Klee“, 1992).

Immediately after the audition listeners were asked to put down their subjective experiences. „Which personal thoughts, emotions, associations does the music evoke inside you? Which bodily reactions do you feel? If possible give your interpretation in one word. What of the music is strange to you, what familiar?“

118 pupils and students (52 male, 60 female, 6 missing values), in the age between 17 and 25 gave their verbal reports.

Steps of interpretation

The first step of the content analysis gives a description of the categories and leads to a transformation of the original text into a sequence of categories.

In a second step contingencies between categories were computed. The contingency coefficients together with the values of approximate significance show the connections in and between the categories on the four different hierarchical levels.

Results

Starting from the main question of how the body is represented in the categories especially in the metaphors and in the music thinking, only some examples of the findings may be reported here:

♦ from the categories: sensation of touch and body as whole, the schemes of container and landscape etc.

♦ from the contingencies: the association of evaluation and familiarity, the connection of gender, messages with the other categories, and also the significant values between body and all other categories will be analysed.

The results give a catalogue and an empirically proofed distribution of image schemes and metaphors specific to music. Also the inner differentiation and intern connection of musical metaphors may appear.
### Figure 1: Categories and frequencies

<table>
<thead>
<tr>
<th>variables</th>
<th>frequencies (percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>persons</td>
<td>118</td>
</tr>
<tr>
<td>school/course</td>
<td>8</td>
</tr>
<tr>
<td>age group</td>
<td>17-23</td>
</tr>
<tr>
<td>gender</td>
<td>male: 52, female: 60, ?; 6</td>
</tr>
<tr>
<td>music</td>
<td>1-4</td>
</tr>
<tr>
<td>1. musical thinking</td>
<td></td>
</tr>
<tr>
<td>1.1 aesthetic evaluation:</td>
<td></td>
</tr>
<tr>
<td>(like it, beautiful) evaluation+</td>
<td>198 (41.9 %)</td>
</tr>
<tr>
<td>vs. (dislike) evaluation-</td>
<td>116 (24.6 %)</td>
</tr>
<tr>
<td>1.2 labelling:</td>
<td></td>
</tr>
<tr>
<td>familiar</td>
<td>267 (56.6)</td>
</tr>
<tr>
<td>vs. unfamiliar</td>
<td>125 (24.6)</td>
</tr>
<tr>
<td>culture, epoch, genre, style</td>
<td>186 (39.4)</td>
</tr>
<tr>
<td>composer, artists</td>
<td>32 (6.8)</td>
</tr>
<tr>
<td>1.3 interpretation/message</td>
<td>313 (66.3)</td>
</tr>
<tr>
<td>1.4 musical elements, structures</td>
<td>102 (21.6)</td>
</tr>
<tr>
<td>1.5 biographic context</td>
<td>62 (13.1)</td>
</tr>
<tr>
<td>1.6 everyday experience</td>
<td>98 (20.8)</td>
</tr>
<tr>
<td>2. language metaphors</td>
<td></td>
</tr>
<tr>
<td>2.1 associations</td>
<td>233 (49.4)</td>
</tr>
<tr>
<td>2.2 emotions</td>
<td>321 (68)</td>
</tr>
<tr>
<td>2.3 space</td>
<td>138 (29.2)</td>
</tr>
<tr>
<td>2.4 time</td>
<td>89 (18.9)</td>
</tr>
<tr>
<td>3. general image schemes</td>
<td></td>
</tr>
<tr>
<td>3.1 container</td>
<td>50 (10.6)</td>
</tr>
<tr>
<td>3.2 building, architecture, landscape</td>
<td>74 (15.7)</td>
</tr>
<tr>
<td>3.3 resonance</td>
<td>53 (11.2)</td>
</tr>
<tr>
<td>3.4 balance</td>
<td>18 (3.8)</td>
</tr>
<tr>
<td>3.5 flowing</td>
<td>27 (5.7)</td>
</tr>
<tr>
<td>3.6 suspense</td>
<td>18 (3.8)</td>
</tr>
<tr>
<td>3.7 attract</td>
<td>21 (4.4)</td>
</tr>
<tr>
<td>vs. repulse</td>
<td>26 (5.5)</td>
</tr>
<tr>
<td>3.8 effects:</td>
<td></td>
</tr>
<tr>
<td>- pleasure</td>
<td>63 (13.3)</td>
</tr>
<tr>
<td>- dreaming</td>
<td>41 (8.7)</td>
</tr>
<tr>
<td>- concentration, reflection</td>
<td>65 (13.8)</td>
</tr>
<tr>
<td>- relaxation, rest</td>
<td>140 (29.7)</td>
</tr>
<tr>
<td>- uneasiness, oppression</td>
<td>79 (16.7)</td>
</tr>
<tr>
<td>- activation</td>
<td>63 (13.3)</td>
</tr>
<tr>
<td>4. bodily sensations</td>
<td></td>
</tr>
<tr>
<td>4.1 sense of touch, other senses</td>
<td>74 (15.7)</td>
</tr>
<tr>
<td>4.2 organism, body as a whole,</td>
<td>285 (60.4)</td>
</tr>
</tbody>
</table>
The music examples show positive as well as negative evaluations, they also possess different grades of familiarity.

The most frequent mentioned categories are: effects (together 451), interpretation/message (313) in the form of emotions (321), associations (233), and bodily sensations (together 359). The in the first position maintained effects nearly at every time are closely bound up with sensations of the body. (These findings support the main hypothesis of this study.) The possible effects may vary by the chosen music examples, but apparently in our every day experience of music we anyway expect such effects. The metaphor of the space is more often mentioned than of time, but very often they are connected quite closely. From the image schemes most often music is compared with a building, an architecture or an landscape. The schemes of resonance and container follow. The other like balance, flowing, suspense and attraction vs. repulse may be of a mere individual importance.

In the computation procedure contingencies were chosen as measure for the connection between two variables, it has a range between 0 and 1. Values reaching from .10000 to .25000 show a relatively strong association between variables. The significance measure has to be near .00000 and should not exceed a value of .02000. The following sample shows the most remarkable values.

**Figure 2: Contingencies inside and between structural levels**

<table>
<thead>
<tr>
<th>Categories</th>
<th>Contingency Coefficient</th>
<th>Approximate Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>like/familiar</td>
<td>.25573</td>
<td>.00000</td>
</tr>
<tr>
<td>like/unfamiliar</td>
<td>.19503</td>
<td>.00002</td>
</tr>
<tr>
<td>familiar/biographical</td>
<td>.23292</td>
<td>.00000</td>
</tr>
<tr>
<td>unfamiliar/associations</td>
<td>.13601</td>
<td>.00286</td>
</tr>
<tr>
<td>gender/message</td>
<td>.13993</td>
<td>.00897</td>
</tr>
<tr>
<td>gender/associations</td>
<td>.12642</td>
<td>.02164</td>
</tr>
<tr>
<td>gender/emotions</td>
<td>.12287</td>
<td>.02684</td>
</tr>
<tr>
<td>gender/landscape</td>
<td>.11215</td>
<td>.04949</td>
</tr>
<tr>
<td>gender/pleasure</td>
<td>.11696</td>
<td>.03789</td>
</tr>
<tr>
<td>gender/touch</td>
<td>.12677</td>
<td>.02118</td>
</tr>
<tr>
<td>gender/body</td>
<td>.12734</td>
<td>.02044</td>
</tr>
<tr>
<td>message/associations</td>
<td>.23107</td>
<td>.00000</td>
</tr>
<tr>
<td>message/emotions</td>
<td>.38292</td>
<td>.00000</td>
</tr>
<tr>
<td>message/space</td>
<td>.15088</td>
<td>.00000</td>
</tr>
<tr>
<td>message/touch</td>
<td>.10941</td>
<td>.01679</td>
</tr>
<tr>
<td>message/body</td>
<td>.12732</td>
<td>.00529</td>
</tr>
<tr>
<td>emotions/landscape</td>
<td>.14431</td>
<td>.00153</td>
</tr>
<tr>
<td>emotions/relaxation</td>
<td>.19307</td>
<td>.00002</td>
</tr>
<tr>
<td>emotions/uneasiness</td>
<td>.13589</td>
<td>.00298</td>
</tr>
<tr>
<td>space/time</td>
<td>.28510</td>
<td>.00000</td>
</tr>
<tr>
<td>space/flowing</td>
<td>.14110</td>
<td>.00196</td>
</tr>
<tr>
<td>space/suspense</td>
<td>.13819</td>
<td>.00243</td>
</tr>
<tr>
<td>time/landscape</td>
<td>.14801</td>
<td>.00115</td>
</tr>
<tr>
<td>time/balance</td>
<td>.15658</td>
<td>.00057</td>
</tr>
<tr>
<td>time/flowing</td>
<td>.18140</td>
<td>.00006</td>
</tr>
<tr>
<td>body/message</td>
<td>.12732</td>
<td>.00529</td>
</tr>
<tr>
<td>body/container</td>
<td>.12305</td>
<td>.00706</td>
</tr>
<tr>
<td>body/resonance</td>
<td>.20155</td>
<td>.00001</td>
</tr>
<tr>
<td>body/flowing</td>
<td>.10567</td>
<td>.02096</td>
</tr>
<tr>
<td>body/pleasure</td>
<td>.10087</td>
<td>.02762</td>
</tr>
</tbody>
</table>
Discussion

The content analysis of the verbal reports describing the inner experience of selected music pieces lead into two directions:

♦ the everyday emotional experiences of music;
♦ the process of generation of metaphors in unguided phrase production (connection between metaphors, image schemes, bodily experiences etc.).

It may be summarised that the essence of the cognitive process consists in an inner movement from bodily experiences to the mind. Thinking in music uses metaphors of language, which go back to basis image schemes. The sense of touch and the other senses together with the feeling of the body as a holistic organism create the fundament of what we experience in or with music.

Familiarity or strangeness have shown significant effects upon the language process. Also the genders experience music differently.

In their report about strong musical experiences Gabrielsson & Lindström (1994, 131) summarise seven basic categories, which may be applied for music experiences in general (131 ff.). What there is referred to as physical responses are equal to my experiences of the body as a whole, the category of perceptions is equal to my sense of touch and other senses. Also the other categories (cognition, emotion, transcendental/existential aspects and personal developments) coincide, but Gabrielsson & Lindström give them another structure than I found in my content analysis.

The results could be included into a body grounded theory of music experience. Some useful elements of this theory I find in the writings of philosophers, psychologists, and pedagogues (Ernst Cassirer, Susanne K. Langer, Claude Lévi-Strauss, Roland Barthes, Mihaly Csikszentmihalyi, Jean Piaget, Robert Haskell, David J. Elliott - just to give a very subjective selection).

Howard Gardner e.g. concerns the metaphoric capacity in the context of his theory of multiple intelligence (Gardner, 1985). May be, a sort of metaphoric intelligence could be integrated into his set of multiple intelligence. Gardner presents his forms of multiple intelligence without showing their relevance and inner connections. So the kinaesthetic intelligence and the space intelligence, although in an essential way correlated and possibly founding the other forms of intelligence, appear side by side with the others. But it could be expected that a hierarchy exists. Especially the socialisation of human intelligence by symbols would have a somehow prominent position within the cognitive potentials, it should be positioned at the top of the hierarchy of the cognitive achievements.

My personal experience in the performing of this qualitative experiment is somehow ambivalent because of two reasons: 1) In the analytic process of data interpretation I sometimes was forced to change, to add, to delete, to modify my modes of reflection. So content analysis permanently is proceeding and never will be completely closed. 2) In spite of the computational work a lot of subjectivity remains. It concerns selection of music examples, selection of the probands, situation of listening, questions to the probands, person of the researcher especially whilst the process of interpretation.
I finish with a more general statement to the qualitative experiment I explained. I refer to the intense discussions about qualitative research in music education (see Reimer, 1996). In Elliot W. Eisner's essay two statements may enlighten our debate:

1) "Human knowledge is a constructed form of experience and, therefore, a reflection of mind as well as nature. Knowledge is made, not merely discovered." 2) "The form through which humans represent their conception of the world not only have a major influence on what humans are able to say, they also influence what they are able to see." (Eisner, 1996, p. 11) If I look at the mainstream research in music cognition of the last decade and if I ask, what are the experimental studies able to see from the music and from the individuals using the music in their everyday life, I feel the necessity to expand the experimental methodologies. Qualitative research methods could be able to bridge the gap between empirical data and everyday experience of music. May I provocingly ask: What is the benefit of a sly experimental design, a very scrutinised data collecting, and all the possible statistical procedures, if the connection to the music, the people and the real life has gone lost?

References


Keywords

music reception, verbal reports, content analysis, familiarity, gender, metaphor, cognition
Varieties of Music Experience

Alf Gabrielsson
Department of Psychology, Uppsala University, Sweden

Few things regarding music can be more important to study than how human beings experience music. Whatever function a piece of music may have in a given situation, it is experienced and responded to in a way that depends on an interaction between factors in the music, in the individual and in the situation where it all takes place. Two people listening to the same music in the same situation will never describe their experience in exactly the same way. The descriptions can differ widely, as we will see in some examples presented later. And one and the same person listening to the same piece of music in two different situations may perceive the music very differently in these situations.

This is no news whatsoever, but yet very little has been made in research on music experience to face this complicated question. Looking for factors that may account for the varieties of music experience, most researchers turn - quite naturally - to the music and try to disentangle various factors in its structure that may be considered as crucial factors behind the experience. Numerous descriptions in textbooks, music essays, program notes etc suggest that a given piece of music has some definite expressive properties and is, or even should be, experienced in a certain way, at least by an "ideal" listener. Anyone among us can give an example of such descriptions, which may in fact be understood, by many people, as normative for how the music should be approached and understood. For a "non-ideal" listener, unknowing of such descriptions, the experience may be something quite else - to him or her such descriptions may appear strange or quite ununderstandable in relation to what they experience.

Much as music itself can vary in its structure, individuals vary - even more - among themselves, both generally and with regard to factors that may affect music experience. This may include our whole life history, determined by genetic and environmental factors in extremely complex interaction, resulting in differences regarding bodily constitution, physiological functions, brain workings, abilities, temperament, attitudes, preferences, and much others. Likewise the situations in which the music happens to be experienced represent an enormous variety, needless to specify further here. Considering this multitude of factors within each of the three domains discussed - music, individual, situation - and the endless, always changing combinations of them, it would be surprising, indeed, to find that a piece of music be experienced in the same way across all these combinations. (I have sometimes thought that a reasonable criterion of "good" music could be that it is appreciated regardless of whatever such combination occurs. But probably no music would stand that test.)

This is a short background to explain the title of this paper, "Varieties of music experience". I am certainly anxious to investigate both commonalities and varieties of music experience, but here I concentrate on the varieties. I think it is important to realize the varieties in order not to unduly simplify or homogenize music experience. The problems I touch upon extend far beyond what can be treated in a short paper like this. What follows is simply a progress report of on-going studies of music experience, retrospective and in real time, both illustrating the varieties of music experience.
I have earlier reported on research concerning especially strong experiences of music (Gabrielsson, 1989, 1991; Gabrielsson & Lindström, 1993, 1995). People have been asked to describe, in their own words and in as much detail as they could, the strongest experience of music that they ever had. They also answered questions concerning the background to the experience; moreover, if they had experienced the music earlier and/or later in life and if the experience then was similar or not; if they had had similar strong experiences in other contexts than with music; etc. They also completed a questionnaire containing statements on strong experiences of music by indicating their degree of agreement with each of these statements. The complete material comprises about 1000 descriptions provided by about 800 persons. Most of it still remains to be reported in enough detail.

The descriptions have been subjected to content analysis, in several rounds, and the resulting classification comprises seven main categories, within which there are two further levels. Conventional psychological terminology is used as far as possible. The seven main categories have well-known and very general labels: General characteristics, Physical responses, Perception, Cognition, Emotion, Existential and transcendental aspects, and Personal development. Only a short summary is given here to exemplify the varieties of the experiences.

The Physical responses in connection with strong experiences are mostly well-known, such as shivers along the spine, goose flesh, tears, changes in breathing and heart rate, even dizziness, and many others. In this category we also include quasi-physical responses such as to feel weightless, soar above the audience or feel carried away by the music. Perception includes at least four modalities: auditory, visual, tactile and kinesthetic, moreover a category representing intensified perception. Cognition involves sub-categories such as altered attitude (e.g., to feel open-minded, abandon analytical thinking, be intensely concentrated); a special relation to the music (e.g., to feel surrounded by the music or be in unison with the music); altered experience of body, time and space; appearance of memories, associations, images; and a cognition-emotion transitional category (e.g., to experience wonder, feel directly addressed, perfection, "this cannot be true"). Emotion includes a variety of "positive" feelings such as joy, happiness, blessedness and so on in ever increasing degree up to euphoria, as well as feelings of safety and warmth, harmony and incredible peace. However, some strong experiences were associated with "negative" feelings, such as intense grief, tension, fear, even horror and panic.

The existential and transcendental aspects of the experience may possibly be described in terms of cognition and emotion but represent phenomena of such dignity that we found it necessary to ascribe them a category of their own. They include thoughts, and often changes, regarding one's life style, relations to others and to life in general. There are plenty of religious experiences, and/or experience of other kinds of existence, sometimes cosmic experiences. The last main category, personal development, gives examples of healing experience, catharsis, new insights concerning oneself and others, strong feelings of confirmation, increased self-esteem, and sense of community with others.

This classification is based upon hundreds of descriptions. Each single description usually contains only a small fraction of the material cited above. Each of them means a unique event, and a mere enumeration of categories as above cannot do justice neither to the variety nor to the uniqueness of the reported experiences. It is simply necessary to read them as they are told - several examples are cited in Gabrielsson & Lindström (1993, 1995) - and it is presently hard to see any other method that can reach so deeply into music experience as using people's own descriptions, although subject to possible distortions of memory as well as to limitations of
verbal skill. They should preferably be studied in combination with results obtained by other methods, verbal or not, in order to possibly reach some consensus.

Immediate descriptions
Of course, strong experiences like these are exceptional (some persons claimed that it had occurred to them only once in their life) and thus not very representative for music experience in general (if there exists anything like that). In a series of on-going studies I therefore have people unexpectedly listen to various pieces of music and ask them to describe their experience. Their descriptions are thus made in real time while the music is sounding and some minute(s) after. For practical reasons the listeners are mostly students, but I also take the opportunity to conduct this type of investigation in connection with courses for audiences of more mixed character. The selected pieces of music are short, usually a few minutes and unknown to practically all participants (this is checked by questions afterwards) to avoid any associations to earlier listenings or events.

In comparison with the descriptions of strong experiences, the content analysis of these reports show both similarities and differences. Examples belonging to several of the above categories appear now as well. There are various physical reactions, although not at all as strong as those described earlier (of course, the situation as such - sitting together with many other people - has an inhibiting influence). Perceptual phenomena mostly refer to auditory, tactile and kinesthetic sensations; visual impressions are absent since the music is not presented live but is recorded. Phenomena belonging to the Cognition and Emotion categories are frequent (some examples follow later) but not as intense as with the strong experiences. However, there are practically no phenomena mentioned that can be referred to the last two categories, that is, no reports about existential or transcendental phenomena, nor about personal development. One can conclude from this, then, that these latter two aspects distinguish strong experiences of music from more ordinary, common-day experiences of music. This should be no surprise, given the circumstances in which these studies are conducted.

Even with this restriction, however, the variety of music experiences among a number of people - sitting in the same room and listening to the same music - is striking. Both myself and my students wonder at the variety of descriptions given under these restricted circumstances. It is perhaps even more striking when considering that most listeners are students in music psychology as well as musicians and music teachers, who are possibly more homogeneous in musical respects than other groups of "ordinary" listeners.

Only one example is given here to illustrate the variety. The music in question is a short piece of piano music, a little more than two minutes, composed some years ago. It begins with only one voice and after a while a second voice enters. I refrain from further structural descriptions to let the listeners speak instead. Three groups of persons have listened to the piece. It was played three times after each other with a short break between the repetitions. During the first presentation the subjects were supposed to just listen, and then to write down their impressions during the second and third listenings and, if wanted, some minute(s) after that. They further answered questions regarding familiarity or not with the music, its position on a simplicity - complexity continuum, how they liked the music, and if they had used any special strategies for listening and reporting etc.

Listening to this piece, you may now make up your mind concerning how you would describe your experience of the piece. You can then compare it with some examples given by listeners in these on-going studies. (The piece is played.)
All descriptions in the following are, of course, translated from Swedish. It is sometimes hard to get a quite adequate translation. The first example goes like this:

"It begins monophonic, atonal, always played with sustain pedal. Extends from the middle octave and two octaves upwards. It goes towards a kind of peak and then falls back. Difficult; impossible to memorize; the groups are perceived as a number of blocks, about 20. Possibly a hard-defined building-of-blocks that is varied. Impossible to group rhythmically."

(Listener's comments: "Pretty interesting; exciting because it is hard to place stylistically. Funny piano music."

This description obviously concentrates on structural features. Compare with the following one: "It is like a bird who starts and flies for the first time. At first hesitating and with gawky wing-strokes. It feels a bit insecure, but then it gets courage and soars, flies and dives. Sometimes it gets a little insecure, but the feeling of freedom gives it courage and strength to fly even to China. - Shimmering and a little hard like icicles dripping from the roof on a warm winter day."

(Listener's comments: "First I listened and got acquainted with the music and which mood it conveyed to me. Then I wrote what I saw in front of myself."

There are at least two different images in this description.)

Next one: "An uneasiness trying to find foothold, a point to repose, a single naked tone to stand upon. An echo sufficient in itself. But scrolls itself further, gets intensified, decays but never, never stops as long as it carries on.... despite the continuous movement, despite repetition and continuity in its "schema", every single tone shoots in like a spear, with pain. Sharpened steel, clinically pure; still the sharpness prevents, makes the depth bearable."

(Listener's comments: "I cut off all 'musts', it becomes almost always an emotional listening. Hard to describe the feeling in words since it in fact is feeling (most)."

And still another: "The music sounds as if taken from a television program that deals with treatment of grief or tragic, depressive events, for instance, treatment of grief at somebody's death, badly treated children in an orphanage or something like that. It may also be someone reporting a traumatic event, for instance, being victimized. The music gives me an impression, at the same time, of thoughtfulness and melancholy. It is in some way beautiful and vulnerable and melancholic." (No comments given.)

It is tempting to continue with more examples but this small sample out of 70 persons' descriptions may in this context be enough to illustrate the variety of experiences of this piece. In fact no two persons came very close in their descriptions despite that they listened to the same music and in a given physical situation. This points to the multi-faceted influences of factors related to individual beings, from their past as well as in the present. The first example reminds of what often is described as analytical or objective listening with no traces of emotional content. Without going into a deeper analysis here, the following three reports represent variants of descriptions in cognitive and emotional wordings but are very different among themselves. The listeners' own comments give some clues about the attitude towards the task. As seen above, one of them explicitly pointed out the difficulty to verbally describe the experience, and the corresponding description may, for a critical reader not knowing the music, appear vague and hard to understand in its use of various metaphors to try to catch the elusive experience.

It is inevitable to reflect upon the use of language for describing music experience. There are well-known pitfalls: individual differences in vocabulary, verbal skill, willingness to use language etc. A specific question concerns the theoretical distinction between on one hand
describing the music "as such" and on the other hand describing one's own response. In these and other reports some wordings are often ambiguous in this respect: they may refer to either of these alternatives or perhaps to both; the border between them may in reality not be very sharp. If a person says, for example, that the music is "cold" - does it mean that he/she perceives the music as "cold" in itself, or that one gets a feeling of "cold" that one ascribes to the music, or can both alternatives be valid at the same time?

The notion about various fixed types of listeners - for instance, "objective", "associative", "emotional" etc - is nowadays mainly abandoned. It is more reasonable to, like Mursell (1937, p. 218), talk about listening types, that is, various ways of listening that can be used by different individuals as well as by the same individual in different situations. The human mind is flexible and permits us to change, consciously or not, between different strategies in listening. Even admitting this, it has been claimed (e.g., Hedden, 1973) that intellectual/analytical listening and feeling-oriented listening usually not occur together. However, in our listeners' reports it is not unusual to find analytical aspects and feelings reported in close connection, moreover statements concerning perception and other aspects of cognition - as in the following highly condensed report "Ah! Exciting, charged. Asian? Lovely. Pentatonic? Safe. Peaceful. Beautiful, clear sounds. Big green mountains, pure air."

The varieties of human music experience are a fact that must not be concealed in strivings at finding generalities or commonalities. We shall look for the appearance of both commonalities and varieties and try to investigate what factors that influence and distinguish between them.

Acknowledgment
The research reported here was and is supported by The Bank of Sweden Tercentenary Foundation.

References


Lexical Cohesion in Linguistic and Musical Discourse

Christina Anagnostopoulou*
University of Edinburgh
Faculty of Music and Dept of Artificial Intelligence
chrisa@music.ed.ac.uk

Abstract: An analogy is presented between music and linguistic discourse in terms of lexical cohesion. Linguistic discourse and music are two different manifestations of human communication. They are both meaningful and coherent, and both are processes inherently temporal. Lexical cohesion is a semantic relation manifested in the lexical level of language. A short sample of discourse is analysed and compared with a sample music analysis, regarding lexical cohesion. Their similarities are discussed, and the term Musical Cohesion is established.

1 Introduction

Linguistic discourse analysis, namely the analysis of text, and music analysis have developed in parallel and have studied similar phenomena. However, no systematic comparison of both techniques and their potential results exists to date. In this paper, linguistic discourse and music are compared with respect to cohesion. First, a linguistic analysis of a short sample of text is performed to demonstrate the concept of lexical cohesion. The musical counter-part follows, with an extended and modified paradigmatic analysis. The two are juxtaposed, making the similarities explicit, and the term Musical Cohesion is established. Finally, the property of linguistic cohesion contributing significantly to coherence and intelligibility is discussed in terms of music.

The analogy between language and music is here investigated from a perspective which is different to various influential existing approaches in three respects. Firstly, the focus is on discourse rather than sentence level. This seems appropriate since both music and linguistic discourse are instances of human communication, carry some kind of meaning, are intelligible, and above all both are inherently temporal. Secondly, a semantic rather than a syntactic relation is examined, namely cohesion,

*I would like to thank Hugh Trappes-Lomax and Raymond Monelle for helpful discussions
which makes it possible to address a semantic level in music, and finally, the investigation focuses on the surface level by looking at associative features, without making any claims concerning underlying structure.

2 The Discourse Analysis Background

Discourse\(^1\) is any linguistic passage that forms a semantic unity, spoken or written, of whatever length or form. The major and necessary factor that causes a linguistic passage to be a text rather than an arbitrary string of sentences is cohesion. Cohesion occurs when some element in the discourse either presupposes the existence of another for it to be interpreted, or is semantically linked to one. For example, in the following beginning of a discourse: “A thermodynamics professor had written a take home exam for his graduate students (...)”, the interpretation of his presupposes the existence of thermodynamics professor. This type of cohesion is called reference.

The type of cohesion discussed here is lexical cohesion, a semantic property manifested on the lexical level by the use of specific words that are either identical or semantically close. It can be divided into reiteration and collocation.

Reiteration includes the exact repetition of a word, a synonym, super-ordinate, or general word, for example, in “I turned to the ascent of the peak. The ascent is perfectly easy”, the word ascent is repeated. Instead of its second occurrence, one could have also used the climb, task, thing as examples of reiteration ([4], p.279).

Collocation is manifested by the use of words that are semantically related in some more distant way, but can still be thought of as belonging to the same semantic network. Examples are the pairs exothermic-endothermic and answer-proof in “Is hell exothermic or endothermic? Support your answer with a proof.”

The above examples demonstrate only pairs of related words, where in fact there can be whole strings, named cohesive chains. In “As for souls entering hell, let’s look at the different religions that exist today (...)”, the chain is souls-hell-religions.

Lexical cohesive chains are sets of words classified together according to semantic closeness or similarity. There can be several chains in a text, running in parallel, and interweaving. They can be global (during the whole discourse) or local (for a part of it).

2.1 An example analysis

In the following text, five cohesive chains are displayed by using different font styles. Reference is also noted (by an asterisk), but only when the presupposed item belongs to one of the lexical chains. Reiteration and collocation are not distinguished.

\(^1\)The approach described here follows [4], which is considered to be the standard and most well-accepted account on cohesion. Note that the terms Discourse and Text are used interchangeably.
Soon her eye fell on a little glass box that was lying under the table: she opened it, and found in it a very small cake, on which the words "eat me" were beautifully marked in currants. "Well, I'll eat it," said Alice, "and if it makes me larger, I can reach the key; and if it makes me smaller, I can creep under the door; so either way I'll get into the garden, and I don't care which happens!"

She ate a little bit, and said anxiously to herself, "Which way? Which way?" holding her hand on the top of her head to feel which way it was growing, and she was quite surprised to find that she remained the same size: to be sure, this generally happens when one eats cake, but Alice had got so much into the way of expecting nothing but out-of-the-way things to happen, that it seemed quite dull and stupid for life to go on in the common way.

So she set to work, and very soon finished off the cake.

(text quoted in [4], p.319).

Figure 1 shows how the above chains (together with their references) unfold through time: the x-axis represents the word number of the discourse, and the level on the y-axis represents the different chains: The first chain (her eye, Alice, her hand, her head, Alice) is shown at level 5, the second (cake, eat, currants, eat, ate, eats, cake, cake) at level 4, the third (larger, smaller, growing, size, things) at level 3, the fourth (way, way, way, into-the-way, out-of-the-way, in-the-common-way) at level 2, and the last (dull, stupid) at level 1. The chains alternate, and some of them are global like "Alice" and "cake", and some local, like "dull".

![Figure 1: Cohesive Chains from linguistic discourse](image-url)
3 The Music Analysis

Repetition, variation and transformation have been studied extensively in music. Prime examples are paradigmatic [6] and motivic analysis, and various other classifications of musical material have been carried out, for example [3, 2, 1].

The method of analysis chosen here is a type of paradigmatic analysis, carried out using a computational model of paradigmatic analysis [1]. The analysed piece is Debussy's *Syrinx* for solo flute. The results, some of which are shown here, were very close to Nattiez' second paradigmatic analysis of the same piece [6]. In order to obtain a classification, each musical segment is described as a list of features. These features are chosen by the analyst, and they can be any musical property, for example concerning melodic shape, rhythm, and whatever the analyst chooses to be his criteria for classification. The segments (described as lists of features) are classified by the paradigmatic analysis algorithm. The categorisation of the segments is hierarchical: there can be many levels, and categories can be divided into subcategories. Also, segments that are not repeated are left out.

Figure 2 shows three sample classes: classes A and B are global, whereas C is local. Most of these segments (apart from the ones in class C) are repeated throughout the piece, but here they are included only once in each class².

![Figure 2: Some of the classes from an analysis of Debussy's *Syrinx*.](image)

²Due to space limitations, the whole analysis is not included here. The method of paradigmatic analysis is considered familiar.
4 Comparison

From the above analyses, two points can be made regarding the similarities between discourse and music.

- Classes of objects can be observed where the objects share common properties and are classified together because of their similarity. There exist various classes with contrasting material which nonetheless can also share certain properties, and form a hyper-class at a higher level.

- These classes are distributed over time. Some classes can be local (like the "dull" chain and the C class above), and some can be global, like "Alice" or A. Classes alternate, following no specific rule. However, one could observe some patterns of sequences, for example occurrence of class 1 might always be followed by occurrence of class 2.

4.1 Musical Cohesion

The formation of classes in both linguistic discourse and music depends on similarity: repetition and variation. In discourse analysis the effect created by these principles is called cohesion. In music, since there is no term for such an effect, the linguistic term can be adopted as musical cohesion.

It is important to note that cohesion is a relation that appears in the text itself, it is visible or audible, and can be brought out and studied. Therefore, it is a relation that appears at the Neutral Level of Discourse. Similarly, musical cohesion appears on the Neutral Level of music. It can be studied objectively and formally.

It has been pointed out that cohesion is a semantic relation, manifested, in the case of lexical cohesion, at the lexical level. In the same way, musical cohesion is not a relation of the equivalent on the phonological level in language, but a semantic relation, manifested in sounds. The similarity criteria for classification were musical properties extracted from the musical segments (and not the segments themselves). These properties can be linked formally into a semantic web.

4.2 Discussion

Coherence is an attribute a text possesses on the aesthetical level, that is in its perception: it is a reaction that we have to a text that it "hangs together". Cohesion supports coherence, although it is not the only factor contributing to it. It could be argued that musical cohesion is a major factor contributing to coherence, although the degree to which this holds might vary in comparison to language.

3 According to Molino's distinction of the three levels, Neutral, poetic, aesthetic, [5].
This issue, and the investigation of other significant factors contributing to coherence (like context) are beyond the scope of the current paper, but are a major direction for future work. Moreover, it would be interesting to investigate how other types of cohesion can be related to music, and in general to explore what discourse analysis has to offer for music analysis. However, as with any parallelism between language and music, although there are self-evident similarities, one should not neglect their differences.

References


Music and Neuropsychology
Timbre perception and sound duration: an ERP study

Tervaniemi, M.¹, Winkler, I.¹,², and Näätänen, R.¹

¹Cognitive Brain Research Unit, Department of Psychology, University of Helsinki, Helsinki, Finland
e-mail mari.tervaniemi@helsinki.fi

²Department of Psychophysiology, Institute for Psychology, Hungarian Academy of Sciences, Budapest, Hungary

Introduction

EEG and ERPs. The neural basis of auditory perception can be noninvasively investigated by recording electric brain activity (electroencephalogram, EEG) during auditory stimulation. By averaging across several hundreds of EEG epochs following the presentation of stimuli, neural events time-locked to these auditory stimuli can be isolated from the ongoing EEG activity. These electric brain events are termed event-related potentials (ERPs). ERPs can be conceptualized as a sequence of components [14] which have different neural generators and reflect different psychological functions [9].

MMN. An ERP component termed the mismatch negativity (MMN) reflects the traces of auditory sensory memory [12, 13, 16]. The MMN is elicited by discriminable changes in sound parameters such as frequency, intensity, and duration when a tone having this deviating parameter is infrequently presented among an otherwise homogeneous tone sequence [for frequency MMN, e.g., see 18, 21, 24]. However, the MMN is elicited only if the inter-stimulus interval between successive tones is shorter than 10 seconds [17]. The time span of the MMN corresponds to that of the second, longer phase of auditory sensory memory [5]. The MMN reflects the discrepancy between the parameters of the incoming stimulus and preceding stimuli which have been pre-attentively encoded in the cortical memory traces [1, 13].

Pre-attentive vs. attentive processes of auditory perception. Voluntary discrimination of stimulus change is not a necessary prerequisite of MMN elicitation. MMN is elicited even when the subject is performing a task unrelated to the auditory stimuli like reading a book or playing a computer game [2]. The MMN amplitude and latency index the perceptual accuracy of a stimulus change as indicated by the correlation between these MMN measures and the performance in active discrimination tasks: MMN latency is shortened and amplitude enhanced in parallel with increased hit rates and shortened reaction times observed when the same stimuli are to be discriminated [15, 24]. Moreover, the subjects' good performance in the pitch-discrimination part of the Seashore musicality test [11] and also in a more cognitively oriented musicality test developed by Karma [10] was accompanied by an MMN amplitude MMN enhancement recorded in passive (e.g., reading) situations [22]. Therefore, the MMN
can be used to determine the accuracy by which the central nervous system represents auditory sensory information. It should be noted that MMN measures are not contaminated by individual differences in the subject's decision and response strategy.

Previous research has concentrated on pure tones and described several parameters whose infrequent changes elicit an MMN. However, a lot less is known about sounds having a rich timbre. The first experiment of the present series established that the memory traces involved in the discriminative MMN process represent sounds in terms of the perceived pitch rather than of the spectral components [28]. In this experiment, MMN was elicited by an infrequent deviant sound differing only in its virtual pitch [20] but not in its spectral contents from the set of standard sounds. This separation of the virtual and spectral pitch was achieved by employing complex missing-fundamental tones [3]. A subsequent study revealed that resolving the missing-fundamental pitch requires relatively long (>150 ms) acoustic samples during the pre-attentive and attentive processing [27].

The aims of the present study. The present experiment wished to determine whether the MMN is elicited by infrequent changes in the spectral component of sound timbre [7] and whether an accurate representation of timbre requires as long stimuli as that needed to establish the missing-fundamental pitch.

Methods

Stimuli. The majority of the auditory stimuli, the standards, were harmonically rich sounds which consisted of three equiloud harmonic partials (600–2400 Hz) of the 300-Hz fundamental tone. Altogether 9 different harmonically rich stimuli were presented, each with p=0.1. The 300 Hz fundamental was not present in any of the standard sounds, however these tones had a common fundamental frequency (300 Hz) and, therefore, were perceived as having a virtual pitch corresponding to 300 Hz. These harmonically rich stimuli were occasionally replaced by a 300-Hz sinusoidal tone, termed the deviants (p=0.1). The working hypothesis was that if the spectrally rich structure of standard stimuli was represented by sensory memory, then the deviant pure tone would elicit the MMN. To investigate the effect of stimulus duration on the MMN elicitation, two stimulus durations, 150 and 500 ms, were employed in separate experiments (offset-to-onset ISIs being 350 and 400 ms, respectively).

Subjects and Procedure. Ten subjects (age: 20-32 years, 6 females) participated in the experiment employing sounds of 150 ms in duration and 12 subjects (age: 19-27 years, 8 females) participated in the experiment with 500 ms stimulus duration. Subjects were not familiar with the purpose of the study and gave informed consent after the nature of the experiment was explained to them. Subjects were not selected on the basis of their musical education or training. During the EEG recordings, the stimuli were presented via headphones to the subjects' right ear. They were instructed to play a computer game and to ignore the auditory stimulation.
**EEG Recording and Analysis.** The EEG was recorded (0.1–100 Hz band pass; 250 Hz
digitation rate) with 10 electrodes attached to the scalp along the midline (Fpz, Fz, Cz, Pz)
and the coronal arcs connecting the left and right mastoids (Lm and Rm, respectively) via Fz
(Lm, L2, L1, (Fz), R1, R2, Rm). The reference electrode was attached to the nose. The EOG
was monitored from Fpz (vertical eye movements) and at the outer canthus of the right eye
(horizontal eye movements). The EEG was off-line epoched (-50–400 ms from stimulus
onset), averaged separately for standard and deviant stimuli, and filtered below 30 Hz.
Subtraction curves (the average ERP to standard tones subtracted from the deviant-stimulus
response) were calculated to estimate the MMN component. The MMN amplitude was
quantified from the individual subjects’ subtraction curves as the mean amplitude during a 40-
ms time window which was centered at the most negative peak detected from the grand-
average subtraction curve.

**Results**

As ERP data plotted in Figure 1 indicate, the MMN was elicited by infrequent sinusoidal
tones which were presented among harmonically rich tones with both stimulus durations 150
and 500 ms although the perceived pitch in both cases equalled to 300 Hz (150-ms duration
experiment: F(1,9)=55.01; 500-ms duration experiment: F(1,11)=34.85, p< 0.001; paired t-test
for standard vs. deviant ERPs). With 150 ms sounds, the MMN peaked, on the average, at 116
ms from deviant-stimulus onset and with 500 ms sounds, at 144 ms.

![150 ms stimulus duration](image1)

![500 ms stimulus duration](image2)

**Figure 1.** The ERPs elicited by infrequent sinusoidal tones which were presented among
harmonically rich tones with stimulus durations 150 (left) and 500 ms (right) (recorded at Fz
electrode). The shading indicates the MMN amplitude. The scale bars (middle) illustrate the
mean MMN amplitude which was significant with both sound durations.
Discussion

The present results suggest that sensory memory encodes the timbre of auditory information accurately even on the basis of relatively short sound samples. As subjects, being engaged in a visual task, did not attend the auditory stimuli one can conclude that the encoding of the timbre quality is a pre-attentive function of the central auditory system.

As reviewed in Introduction, Winkler et al. [27] found that extracting the virtual pitch of missing-fundamental tones requires longer tone duration than 150 ms. The present results, however, showed that this stimulus duration was sufficient for discriminating sounds of widely differing timbre. Therefore the present results support the notion that pitch and timbre are determined by separate processes in the human auditory system [e.g., 19].

The present result, showing the precedence of timbre extraction to pitch extraction, ties in well with the ecological point of view: In natural acoustical environments, different sound sources most probably also have different characteristic sound spectra. Consequently, some fast pre-attentive process [c.f., 4] must segregate sound sources on the basis of timbre [6, 8].

Still one further aspect of the present results merits consideration. Although the exact timbre of the 9 standard sounds was different, the MMN elicitation by the deviant pure tones suggests that the processes underlying MMN grouped these sounds together, contrasting them with the qualitatively different deviant timbre. This is because if all 9 standard sounds were treated as separate items, then the pure tone could not have become “deviant”, having the same probability as any of the individual standard sounds, whereas, as already reviewed, MMN is elicited by infrequent changes only on the basis of some detected constancy. There are two possible explanations of why the timbre of the standard sounds was regarded as “equal” by the processes underlying the MMN elicitation. First, the difference between the individual timbres of the standard sounds was small enough for the auditory system to take it as the “natural” variability within the same source. Previous results [26] show that the MMN system tolerates some variability in the intensity of the standard ones (an ecologically important feature of this process), provided that the deviant stimulus is sufficiently distinct from the range of standard-stimulus variance. The alternative explanation suggests that, although each harmonically rich standard tone was processed separately since their spectral composition differed, the auditory system abstracted their timbres into one unit, contrasting it with the qualitatively different timbre of the sinusoidal deviant. Such a phenomenon has been observed by Tervaniemi et al. [23]. They presented reading subjects with continuously descending Shepard tones. An MMN was elicited by ascending or repetitive tones when they were infrequently presented among descending tones despite the variability in the absolute tone frequencies. The authors concluded that the system underlying MMN pre-attentively formed the “abstract” category of descending pitch change. The present result does not allow one to distinguish between these two alternative explanations. However, the behaviour of the MMN observed in the present study is compatible with the assumed role of the processes underlying MMN in maintaining an accurate model of the acoustic environment [25].
Conclusion

The present ERP evidence indicates that the spectral component of timbre is pre-attentively processed on the basis of relatively short sound samples and subsequently encoded into cortical sensory memory-traces.

Acknowledgements

This work was supported by the Academy of Finland, the Finnish Society of Psychology, and the National Scientific Fund of Hungary (OTKA T022800).

References


Neural Representations of Music
Evoked by Verbally and Musically Based Learning Strategies
An EEG study on cortical auditory activation patterns

Wilfried Gruhn, Musikhochschule Freiburg i.Brsg.
Eckart Altenmüller, Musikhochschule Hannover, Institut für Musikphysiologie und Musikermedizin
Dietrich Parlitz, Musikhochschule Hannover, Institut für Musikphysiologie und Musikermedizin

The inquiry is concerned with neurological reactions to different learning strategies that evoke discernible types of musical representations. Therefore, the focus of this study is given to the characteristics of genuine musical representation itself, especially to the question whether there is a distinctive difference of mental representation corresponding to the procedure by which the particular representation is developed. Then, different strategies of music teaching and learning will be mirrored by changes in cortical activation patterns in terms of an increase or decrease of the intensity of activation and referring to the localisation within the topography of the cortex.

Here, music learning is understood as the process by which mental representations are established and then incrementally altered. Perception and cognition of music are based upon the activation of already developed musical representations. But the way one processes musical information depends of the cognitive level and neural stage of musical representation. For this, we distinguish - corresponding to Bamberger (1991) who has picked up a term of Piaget (1947) - between "figural" and "formal" representation (see Altenmüller & Gruhn 1997; Gruhn & Altenmüller 1996).

This distinction refers to two types of knowledge acquisition. There is an explicit or declarative knowledge by which we develop a verbal representation of the environment. We know about music, and at the same time we know that and what we know. On the other hand, there is an implicit or procedural knowledge by which we know how to process things and objects. In this case, we know how to play a tune, or how a chord or chord progression sounds like. Music can be represented verbally, i.e. by musical terms which represent a symbolic code or transformation of musical sound into verbal designation. But music as such is different from signs and symbols. Music is mentally represented in terms of audiated sound and musically imagined structures. This "mental image" of sound is not verbal, rather it is originally and primarily bound to an embodied procedure: rhythm is bound to movement, pitch to voice production or the production by an instrument. Therefore, our first mental image of a chord or motif or tune is derived from the fingerings and muscle tensions by which the sound is produced. This kind of representation (which is a typical procedural knowledge) is strictly bound to the sequence of time. A tune, then, is represented by a sequence of actions producing that tune, i.e it is represented by a complex setting of concrete figures. Therefore, this kind of representation is called "figural".

After a while of acting that way, the link of producing sound and the sound itself will be strongly reinforced, the movements have been internalized and finally become automatized. But all of a sudden, a cognitive leap takes place when the tune is no longer represented as a sequence of figural actions, but as a formal structure. Then, it becomes independent of a concrete realization. A change of mental representation has taken place. A tune, then, is encoded in a more general, categorical representation which allows to process it by formal operations.
(Piaget). Here, the neural representation must have been changed, and a hierarchically higher (or cortically deeper) level of neural representation has been developed. This type of representation is called "formal" (according to Piaget's developmental stage of "formal operations"). Learning, now, can be described more precisely as the transition from figural to formal representation.

The experiment

In a longitudinal EEG study, the hypothesis should be proved if different strategies of learning which are referred either to the developing of declarative knowledge or to procedural abilities, can be identified by different cortical activation patterns. All subjects (N= 23; 14 female, 8 male; volunteers from local high schools, age 14 - 15) were right handed (according to Oldfield's handedness inventory) and had got comparable musical education in school and similar musical experiences outside school (singing in a choir, instrumental lessons). Additionally, all were tested by Gordon's standardized music audiation test (AMMA) where they ranked relatively highly (means of 64.3 on the percentile rank) with a distribution from 14 to 91.

The goal was to establish a musical representation of an unfamiliar tonality: the dorian mode. This task was chosen because we wanted to assure that nobody had already developed a figural or formal musical representation of that tonal structure even if they played an instrument; they not even had ever heard the name "dorian". The task was to identify the tonality of a played melody. For this, a modified probe-tone technique was employed. Subjects listened to a short four-to-eight-bar melody and were asked to judge whether the tunes "fit" a preceding cadence which was given to stabilize a tonality (major, minor, or dorian). A set of 60 items (30 same, 30 different) presented the following combinations in a randomized order:

<table>
<thead>
<tr>
<th>cadence</th>
<th>melody</th>
<th>number of items</th>
</tr>
</thead>
<tbody>
<tr>
<td>major</td>
<td>major</td>
<td>10</td>
</tr>
<tr>
<td>major</td>
<td>dorian</td>
<td>10</td>
</tr>
<tr>
<td>minor</td>
<td>minor</td>
<td>10</td>
</tr>
<tr>
<td>minor</td>
<td>dorian</td>
<td>10</td>
</tr>
<tr>
<td>dorian</td>
<td>minor</td>
<td>10</td>
</tr>
<tr>
<td>dorian</td>
<td>dorian</td>
<td>10</td>
</tr>
</tbody>
</table>

The crucial combination was minor - dorian and dorian - minor respectively, because it is likely to assume that both tonalities would sound the same to those who had not yet learned clear criteria for differentiating minor and dorian tonality, in other words who had not developed a musical representation of the dorian sound structure.

After a short training phase when the task was introduced and explained to all subjects, the brain activation patterns of the students were measured using 32 electrodes while listening to the set of items. Then, three subgroups were formed. The aim was to develop a genuine musical representation of the dorian mode in one group (L2; N = 9) just by listening, playing, singing, and moving. No verbal explanation was used, no visual aids employed. The participants should become familiar with and develop and stabilize a "mental image" of the characteristics of the dorian sound. On the contrary, a second subgroup (L1; N=8) were encouraged to develop declarative knowledge about the dorian mode, its half steps and whole steps, the sharps and flats etc. All musical examples served as acoustic "illustrations" instead of exercises for experiencing the dorian mode. Finally, a third subgroup functioned as a control group (NL; N=6). Those participants received the same amount of instruction, but the content of lessons had no relation to any kind of tonality.
The learning phase continued for about three months when subjects met once or twice a week for one hour. At the beginning, students of group L2 were always oriented on their fingerings or asked for notation to memorize the patterns and tunes performed during the lessons. But they were asked just to imagine and memorize the different sounds of major, minor and dorian. For this, subjects got cassettes with training patterns to sing and play including harmonic patterns for improvisation. Only if one is able to improvise in minor and dorian just by listening to given patterns, one can assume that a formal representation has developed. But as long as subjects were dependent of their fingerings, they still remained on a figural level of representation. At the end of this first section of learning, all participants were measured once again, using the same set of items. Afterwards only group L2 continued their training. The aim was to transform the figural into a stabilized formal representation (fig. 1).

Figure 1
Design of the experiment

**EEG measurement**
Neural activation, measured in this experiment, is reflected by excitatory post-synaptic potentials in the outer layer of the cortex. These negative potentials can be recorded directly from the scalp using specially developed electrodes electrodes. These potentials must be amplified and can be demonstrated for each position as an activation curve (fig. 2). The 32 positions allow a sufficient spatial distribution to assess specific activation within different brain areas.
The subjects were measured while listening to the tasks; their answers were recorded and stored in the computer. After averaging the individual data ANOVA was performed. For the comparison of the results, learners in each group were separated from non-learners because obviously (shown by results of the test) not all had fully achieved the task. But there were at least three individuals in each learner group (L1, L2) who had learned (L1L; L2L) whereas others did not show any effect in task solving (L1N; L2N).

Results

With respect to solved items, it is salient that the most remarkable increase is related to tasks for which a clear distinction of minor and dorian modes is needed (fig. 3). The neural activation patterns reveal significant differences with respect to the power of activation and the distribution (localization) according to the different learning strategies. At the beginning (pre-test) all groups show the same topography of activation patterns. That is also true for the NL group in the post-test. Here, one can only observe a slight decrease of activity as a consequence of habituation. In the pre-test, all groups show an activation of bifrontal and temporal regions. This might indicate an involvement of primary and secondary auditory and associative auditory areas.

Looking at the patterns of learning groups L1 and L2, one is faced with a much more differentiated situation. Both groups reveal a shift of activation from bifrontal to right frontal areas, which reaches far into the right temporal lobe only in learning group L2. More important is a second focus of activation in the left parieto-occipital lobe. Whereas subjects of group L1 re-
veal less activity in areas of verbal (declarative) knowledge, learners of group L2 surprisingly have an increase in this area. This group is hallmarked by two more characteristics. Firstly, there is a remarkable reduction of the extension of active areas. Consequently, one can observe a concentration within several small areas which causes a scattered activation pattern. This picture is only shown by learners who have developed a genuine musical representation. That indicates a basic change within the code of the represented musical knowledge. This new way of representation originates from figural representations that have become formal.

![Leistungszuwachs der Lösungen](image)

**Fig. 3**
Increase of solved items in different learning groups. The increase of the total can be referred to the most remarkable increase of tasks for which a clear differentiation between minor and dorian modes was needed.

The difference of cortical activation patterns between the first and the second measurement (pretest - posttest) is highly significant ($p = 0.007$) in L1L and significant in L2L ($p = 0.02$), whereas subjects in group NL do not reveal any significant change ($p = 0.2$). More surprisingly, even subjects of L2N perform a remarkable difference in their brain activities at the critical value of $p = 0.06$. Referring to an interelectrode correlation (indicating the degree of connectivity between brain areas) group L2L produces the strongest difference (from pretest $p = 0.38$ to posttest $p = 0.03$).

Special attention should be given to a comparison of those L2-students who's test indicates a successful learning (L2L) with those who's test did not (L2N). Their activation patterns also show remarkable differences. Most obvious is the focus on the left frontal lobe and the nearly entire elimination of posterior cortical areas (parietal and occipital lobes). Interestingly they share this attribute with subjects of group L1L, but separate them from non-learners (NL). This means that even those who have not gained an observable learning attitude, must have developed a change within their mental representation by the way they have learned, which is different from both, the non-learners and the learners of group L1.

To sum up we can state that in general learners activate on their right fronto-temporal lobe, whereas non-learners predominantly focus on the left temporal brain area. Only subjects of L2
who transformed their figural into a formal representation develop a second activation center on the left parieto-occipital lobe which can easily be related to the transition from figural to formal representation. This verifies the assumption that the employed teaching and learning strategy causes different patterns in the topography of cortical stimulus processing. It could be demonstrated that this directly corresponds with the type of representation which is developed and elaborated through the method of music learning.

References
INTRODUCTION

Little is known about the neuronal networks underlying the processing of emotions. In humans, research focused on observations of changes in autonomic reactions that accompany emotion. In animal studies, it could be demonstrated that the neurobiological bases of these autonomic reactions are specialized neuronal circuits located in the subcortical limbic system (LeDoux 1993). However, it is clear that the cerebral cortex and in particular the anterior cortical zones of the brain play an important role in many aspects of human emotional behaviour and experience (Kolb and Taylor 1981). The first notion that cortical processing of emotional behaviour is lateralized in humans came from one of the founders of modern brain research: John Hughlin Jackson (1879) noted that damage to the left hemisphere was more likely to cause severe depression whereas damage to the right hemisphere produced in some patients inadequate indifference or even euphoria. Later, this observation could be verified in neuropsychological studies on large populations of unilaterally brain damaged patients (i.e. Gainotti 1969, for a review see Davidson 1995). It was concluded that under physiological conditions emotional behaviour is organized in a ‘balanced’ manner. According to this ‘equilibrium-theory’, a unilateral lesion produces a preponderance of the emotions processed in the contralateral hemisphere. As a consequence, it was postulated that positive emotions are primarily processed in the left, negative emotions in the right hemisphere.

Music is undoubtedly a powerful tool to elicit emotions and to modify moods. Autonomic responses to music listening, such as ‘chills’, have been studied thoroughly (Jaak Panksepp 1995). But the neuronal basis of emotional responses to music remains unknown. Up to now, no brain activation studies during emotional music processing are available. The present study was designed to fill this gap. The aim was to clarify whether different emotions during music listening, depend on lateralized neuronal networks, especially in the anterior cortical regions. Furthermore, we were interested in the questions whether lateralized brain responses to music depend on sex or the style of the music. Since it is difficult to control for emotions objectively, we restricted our study to the assessment of brain activity related to attitudes and music preferences, assuming that both parameters are closely linked to emotions.

METHODS

Experimental principle: To assess brain activity, a non-invasive EEG-technique was applied. Cortical activation causes an increase in negative field potential at the apical dendrites of cortical pyramidal cells. The local distributions of these surface-negative low frequency DC-
potentials reflect cortical activation patterns. Since these DC-potentials are lower in voltage than the ongoing background EEG, the signal-to-noise ratio has to be enhanced by averaging task-related EEG-activity over 30 to 40 trials. Activation patterns obtained with this method are highly task-specific and intra-individually reproducible (for neurophysiological details of the method, see Altenmüller, 1993).

**Subjects and general procedure:** 32 right-handed students (16 female, 16 male), aged 12-15 years with similar musical and general education were investigated. All gave informed consent to the proceedings during the experiment. They had to listen to 160 short pieces of music and environmental sounds of 14.5-15.5s duration. During listening, DC-EEG-activation patterns were assessed with 32 electrodes over both hemispheres. After each stimulus-presentation, subjects were asked to rate on a five-step scale, whether they liked the music or sounds presented previously or not (1 = I like very much, 2 = I like, 3 = I am undecided, 4 = I do not like, 5 = I do not like at all).

**Stimuli** were selected from four categories of music or sounds respectively. Out of the overall 160 stimuli, 40 stimuli respectively belonged to the repertoire of classical music, to jazz-music, to pop-music and to environmental sounds. During the experiment, stimuli were presented in a random order. Vocal music was not included in the set to avoid confounding with language related brain activation. Furthermore, to rule out recognition-effects, only unknown pieces of music were selected. In order to cover a broad range of emotional responses, stimuli were chosen from a large variety of heterogeneous styles and pre-selected for presumed emotional reactions. It was planned that about half of the stimuli should produce a more negative and half a more positive response. For this purpose, some items from classical and popular music were pretested with a group of students not participating in the EEG-experiment. Furthermore, members of the staff of our institution (n=3) were asked to select the remaining stimuli with respect to variations in emotional reactions. To give an example, contemporary experimental music produced more frequently negative reactions (don’t like) whereas excerpts from viennais classical music was likely to produce positive reactions. Environmental sounds taken from industrial production, such as noises from a circular saw produced negative, sounds of breaking waves of the ocean produced positive reactions.

Stimuli were prepared using high quality CD or disk-recordings as raw-material. In a first step, the music was digitised and the 14.5 to 15.5s sequences were determined using musically ‘logical’ fragments. It was intended that the short sequences should sound as more or less closed musical phrases. Subsequently, the sound pressure level of each stimulus was adjusted to an average (subjective) loudness of approximately 64 dB. To avoid startling the subjects, beginning and end of each stimulus were faded in or out respectively.

**Recording procedures:** Subjects were seated comfortably 2 meters in front of the stereo loudspeakers. Volume was kept constant in all experiments. To prevent artefacts in EEG-recordings, subjects were requested to fix their gaze on a small circle in front of them and to avoid body movements or vocalisations while listening. Subjects listened to the sequences and were asked to communicate orally their ‘emotional’ judgement right after the end of each stimulus. Subjects were instructed ‘not to think too much about it’ and to rely mainly on their ‘feelings’. While listening, the DC-EEG was recorded from 32 electrodes positioned according to the 10/20 system over frontal, central, temporal, and parietal brain areas of both hemispheres using the standardised Electro-Cap (Electro-Cap International, Eaton OH). A linked-earlobe electrode served as a reference. Impedance was reduced below 3 kOhm. Artefacts arising from eye movements or tongue movements were controlled by bipolar electrode montages. To record the very low frequencies of the EEG, a newly developed
commercially available DC-amplifier system (Neuroscan/USA) was used. The frequency band of amplification ranged from DC to 30 Hz. The amplified signals were digitised at a 100 Hz sample rate and stored.

Data Analysis and Statistics: Trials contaminated with artefacts were excluded from further analysis. Due to the high quality of data required, exclusion criteria had to be applied very strictly. In consequence, several data sets contained less than 8 trials in one of the stimulus categories. In these cases the whole data-set was excluded from further statistical analysis. As a result, only the data of 16 subjects (8 male, 8 female) could be analysed statistically.

For data quantification, mean amplitudes of brain activation during listening were calculated for a distinct time-epoch and for each trial. As time-epoch, the seconds 10-12 were chosen. The mean amplitudes of brain activation during this period were related to the mean amplitudes during a baseline period 1 second prior to the beginning of the stimulus. Data on mean amplitudes were subjected to ANOVA. Within-subjects factors included: (a) „emotional categorisation“ (two levels: positive vs. negative), (b) „stimulus“ (four levels: classic, jazz, rock-pop, environmental sounds), (c) „sex“ (two levels: male vs. female), (d) „electrodes“ (32 levels, according to all electrode positions) and (e) „laterality“, i.e. the difference of activity in homologous pairs of electrodes on either hemisphere (11 levels, according to the number of paired electrodes). In order to detect differences in topographic distribution rather than in amplitude, in a second step the data were normalised according to the procedure proposed by McCarthy and Wood (1985) and ANOVA was repeated on these normalised values.

Results

Melody and sound processing produced a widespread bilateral activation over the anterior parts of the cerebral cortex. The activation pattern included prefrontal and frontal areas, furthermore the anterior and the central temporal regions of both hemispheres. When analysing separately the trials categorised as „positive“ (category 1 and 2) and as „negative“ (category 4 and 5), a clear lateralization effect related to the emotional categorisation of the stimulus emerged. ‘Positive’ emotions produced a significantly more pronounced lateralization to the left hemisphere over frontal (F3-F4, p < 0.05) and fronto-temporal brain-regions (F7-F8 and FT7-FT8, p < 0.01). This effect was mainly due to decreased activation of right-fronto-temporal areas when listening to ‘positively’ categorised music. No emotion-related differentiation in lateralization could be detected over posterior brain areas. The emotion related differences in lateralization occurred only when listening to music, but not when listening to non-musical sounds. The style of music did not influence the brain activation pattern.

The sex of the subjects was an important factor with respect to the lateralization pattern. In female subjects, the emotion-related differences in brain activity comprised widespread brain-regions including frontal, fronto-temporal and fronto-central cortical regions. In contrast, in male subjects lateralization differed only over anterior fronto-temporal brain regions (F7/F8), thus including only a very circumscribed cortical area.

In Fig.1, the lateralization-data calculated for the normalised data set and separated for female and male subjects are shown. On the y-axis, the relative preponderance of activity of either hemisphere is depicted. Negative values indicate left-hemispheric, positive values right hemispheric preponderance. On the x-axis, the electrode locations of homologous pairs of
Fig. 1: Results of ANOVA concerning the factors lateralization, electrodes and sex and emotion-related categorisation.

In the figure it is easy to recognise that emotion related differences emerge almost exclusively over frontal and fronto-temporal brain-areas. While negative emotions yield almost bilateral symmetrical activation patterns over anterior brain regions, positive emotions produce a left-hemispheric lateralization. Furthermore, it is evident that the effects are more pronounced in female than in male subjects.

Discussion

The results of the study can be summarised as follows:

1.) There are emotion-related differences in hemispheric lateralization when listening to music. Positive emotions during music processing are lateralized to the left hemisphere.
2.) These differences can be observed exclusively over frontal and fronto-temporal brain areas, but not over posterior brain areas.

3.) The differences are more pronounced and include larger brain areas in female subjects than in male subjects.

The fact that positive emotions produce a lateralization towards the left hemisphere is in line with the results of lesion studies cited above. However, the lack of a clear lateralization towards the right hemisphere in male subjects when listening to negatively categorised music is surprising. Since no converging data from other studies are available, the explanations must remain somewhat speculative. In non-professional musicians, one would expect a more pronounced responsiveness of the right hemisphere for music in general (Altenmüller 1989). Furthermore, using visual stimuli such as face expressions or films, the preponderance of the right hemisphere for negative emotions seems to be very strong (Wittling and Roschmann 1993). It is not very probable that processing of visually induced emotions differs principally from processing of auditory stimuli.

However, there are two arguments which could explain the bias towards left hemispheric preponderance in the lateralization data. The first concerns the design of the experiment: subjects listened only to short sequences of music and had to categorise them afterwards. This categorization process most probably occurred synchronously to listening. It could be that subjects used internal speech or other more analytical and sequential cognitive strategies to deal with this task. The second argument is more related to musical experience. It is very difficult to induce negative emotions with music. Perhaps the male subjects, - when answering ‘I don’t like this at all’ - felt in the deep corners of their souls that this music was not so bad after all. The answer for categorisation could reflect rather the intellectual concepts and prejudices than the real state of the mood. This would mean that at least a portion of the ‘verbally’ negatively categorised stimuli were ‘emotionally’ attributed to the wrong category, explaining that lack of right-hemispheric lateralization in the ‘negative’ category. In this context it is interesting that hemispheric differentiation was more pronounced in females. We do not know, whether the female subjects were less ‘intellectual’ when listening to the music. Possibly, such an open-mindedness and - as a consequence - improved affective resonance incites in females a broader range of emotional experience, resulting in the increased emotion-related variation in laterality-patterns.

References


Jackson JH (1879) On affections of speech from deseases of the brain. Brain 2, 203-222


Musical abilities in Individuals with Williams Syndrome

Daniel J. Levitin  
Interval Research Corporation, Palo Alto, CA USA

Ursula Bellugi  
The Salk Institute for Biological Sciences, La Jolla, CA USA

We report new evidence for the modularity of music perception and cognition from a work in progress on the musical abilities in individuals with Williams Syndrome. Williams syndrome (WS) occurs in 1 out of 20,000 births, and is associated with a specific heart defect, facial features said to be 'pixie-like' (linked to idiopathic hypercalcemia by Black & Bonham-Carter, 1963), auditory hypersensitivity, and a deletion of one copy of a small set of genes including elastin on chromosome 7. Previous research has shown that WS is also characterized by impaired cognitive function (mean IQ=58), poor spatial, quantitative, and reasoning abilities, coupled with excellent face processing and relatively intact language abilities in older individuals (Bellugi, Klima & Wang, 1996; Bihrle, Bellugi, Delis & Marks, 1989; Reilly, Klima & Bellugi, 1990). Because of the dissociation between language and other aspects of cognition, WS provides a compelling argument for the modularity of language.

Anecdotal evidence has suggested that many individuals with Williams syndrome also possess musical abilities. To date, however, there has been little scientific investigation of musical abilities in WS individuals.

For much of its history, experimental psychology has viewed intelligence as somewhat monolithic, and mental retardation as reflected more or less uniform impairment across domains of cognitive functioning. The study of distinct, well-defined, and atypical populations affords the investigation of specific aspects of cognition. Studies of these populations can provide evidence for the debate over the modularity of specific mental functions, and promises to contribute to a better understanding of the underlying neural architecture in cognitive processes. Bellugi has begun a thorough and systematic study of Williams Syndrome in an effort to deepen our understanding of all these issues (e.g., Bellugi et al., 1996; Karmiloff-Smith, Klima, Bellugi, Grant & Baron-Cohen, 1995).

In this preliminary report, we present both qualitative (n=40) and quantitative (n=10) studies of individuals with WS, attending a summer music camp near Tanglewood, Massachusetts in 1996.

Qualitative observations.

In our qualitative "natural setting" observations at the camp, we noticed that the WS individuals had an unusually high degree of engagement with music. Music seemed to be not just a very deep and rich part of their lives, but one that was omnipresent; most of them spent a great proportion of the day singing to themselves or playing instruments, even while walking to the mess hall. They appeared to do this with a good deal less self-consciousness than normal music camp attendees, or normal individuals. When one camper encountered another camper or group of campers involved in a musical activity, no matter how informal, the newcomer would either join in immediately or begin swaying appreciatively in time to the music. We feel the need to emphasize our observation that this consuming involvement with music is unusual in normal populations; one of us (DJL) spent a number of years working as a professional musician and rarely encountered this type of total immersion among musicians - in fact, the type of connectedness to music the WS individuals naturally displayed is often considered a goal among many musicians.
All the more striking was the WS individuals' inability to carry out tasks requiring low level hand-eye coordination or visual-spatial skills. For example, many subjects had difficulty walking, and handled eating utensils with an uncomfortable awkwardness. Yet, when presented with the musical instrument on which they specialize - be it a clarinet, piano, drum set, or guitar - they are able to execute musical passages requiring a much finer degree of control and coordination than that required to successfully walk a flight of stairs or cut their food. There is a decided lack of technical perfection in all these musical performances, but this is more than compensated for by the subjective quality of "soul" they inject into their performances.

One participant ("B") was known among the other campers as a prolific songwriter. To examine B's ability, the experimenter asked him to write a song about a breakfast cereal, specifically "Kellogg's Rice Krispies," a topic which was a complete non-sequitur during the experimental session. To E's surprise, B spontaneously composed a song about Rice Krispies, complete with verse, chorus, and rhyming lyrics. Without further observations, it is not clear how much this apparently new composition differed from others that B has written, but a quick review of some of B's repertoire revealed both a distinct compositional style, and a marked difference among the various tunes.

In a previous study, one of us (UB) demonstrated that WS individuals make extensive use of paralinguistic channels (Reilly et al., 1990) while listening to stories. Specifically, WS individuals used affective prosody and lexically encoded narrative enrichment devices, including many that are specifically listener-oriented as well as being affective. This places WS individuals on the end of a continuum of social cognition anchored at the deficit end by autistic individuals.

The parents we interviewed all reported that their children have an intense connection to music away from camp as well. Many parents reported that their children knew hundreds, if not thousands of songs from popular music records; although this is not unusual among "normal" American teen-agers, it is worth noting as a contrast to the WS individuals' deficits in other cognitive domains. Several parents reported that their children seemed to have unusual abilities in sound identification. One child could reportedly distinguish among a dozen different brands and models of vacuum cleaners based exclusively on the sound of their motors; another child could reportedly identify dozens of cars as they rounded the corner outside of his house based solely on the sound of the engine. These claims have not yet been tested, but suggest that a skill akin to "absolute timbre" may exist in these individuals.

Quantitative Observations.

Our quantitative measure took the form of rhythmic tests conducted on WS subjects (8 - 24 years old). The experimenter presented each subject with a series of increasingly complex rhythms, clapped on the hands, and asked the subjects to repeat them. The dependent variable was their percent correct score.

Without being explicitly told to do so, the subjects very naturally began to clap back the demonstrations in perfect time, without missing a beat. That is, the subjects interpreted the examples as forming part of a larger musical set; the subjects seemed to assume that there was an implied time signature and tempo, and they responded to the "first measure" of music played by the experimenter in time for the downbeat (or in some cases pickups) to the "second measure." Moreover, the subjects revealed a remarkable ability to track changes in rhythmic pulse, including changes to swing time, straight eighths, triplets, sixteenths, syncopations, and so on.

In roughly 2/3 of the cases, the subjects played back the rhythms correctly (regardless of difficulty). They were wrong on roughly 1/3 of the trials, although their incorrect responses were not necessarily to the more difficult rhythms, raising the possibility that lack of attentional control was responsible for their errors, rather than the rhythmic complexity of the example presented.
The most interesting observation was that in roughly 1/2 of those cases in which the WS subject clapped an incorrect rhythm, the subject altered the original rhythm in a fashion that could best be described as "rhythmic completions." In other words, the pattern they played back on these occasions could be considered part of a "call-and-response" rhythm pair, such as are shared by jazz musicians "trading one bar phrases." To put it another way, the subjects seemed to be making music out their response, rather than slavishly mimicking the experimenter. Two independent data coders who knew nothing about these data or this hypothesis, and themselves professional musicians, confirmed this "rhythmic completion" effect.

We used a set of mental age and sex-matched control subjects, aged 5 -7, recruited from the Palo Alto community, and all of whom were actively involved in music and music lessons. The control subjects did not score significantly different from the WS subjects on the rhythm tests, although their productions were characterized by a lack of the "rhythmic completion" effect.

Conclusions.

We have described evidence for the modularity of musical ability in individuals with Williams' Syndrome (WS). Despite substantial impairments in virtually every cognitive domain, WS individuals maintain relatively intact language and music abilities. We saw no individuals who had intact musical abilities but impaired language function, supporting the conjectures of previous researchers that language and music may have evolved from, and involve, common subsystems.

References.


THE FETAL MUSICAL ENVIRONMENT

R.M. Abrams, S.K. Griffiths, X. Huang, J. Sain,
G. Langford, and K. J. Gerhardt

Depts. of Obstetrics and Gynecology, Communication
Processes and Disorders, and Music
University of Florida
Gainesville, FL 32610, USA

INTRODUCTION.

The intrauterine sound environment is both rich and varied. Analysis of human heart rate responses in a variety of conditions suggest that fetuses perceive some sounds, habituate to them, and on the basis of early postnatal behaviors, demonstrate some learning of acoustic cues heard prenatally. The degree to which auditory experiences shape fetal behavior or modify fetal brain maturation in positive or negative ways, depends in great measure on the characteristics of the sounds reaching the fetal inner ear. These characteristics are defined by the durations and sound pressure levels of the sounds created by the mother-to-be as well as by sounds present outside the maternal body. The spectral features of the sound, the route of stimulation to the fetal inner ear, and the developmental stage of the cochlea and central nervous system are also important considerations.

Much of what we know about acoustic and vibration transmission characteristics and fetal responses comes from experiments in fetal sheep (2,3,7). Similarities in body weight and abdominal composition and dimensions obviate the need for scaling factors. As in humans, the auditory system in sheep is functional well before birth. Investigators have concentrated mainly upon pure tones and noise to map intrauterine attenuation patterns (3), to determine the importance of bone conduction (4) and to exploit vibroacoustic stimulation as a test for fetal responsiveness (9).

Less is known about the characteristics and effects of complex sounds, including those represented by music. The idea that prenatal musical experiences somehow improve the appreciation and performance of music postnatally persists in the popular press. Mothers-to-be have been encouraged to sing to their unborn babies. Specific recorded music is believed by some to be especially enhancing for neural development. While there are likely general features of music, including melody and intonation (8) that can be perceived by the fetus, and are possibly valuable in learning, the known attenuation characteristics of higher frequency sounds (3) would probably render the identification of some specific instruments difficult. The timbral features, by which we identify instruments, would almost certainly be compromised by the filtering characteristics of the abdominal contents.
In the present experiments, we measured the spectral features of tones produced by a synthesizer and by a musician. Sound pressure levels (SPL) in air (extraabdominal) were contrasted with intraabdominal SPL measurements in order to assess changes in spectral content resulting from transmission through tissues.

METHODS
Guidelines for the care and use of animals approved by the University of Florida were followed. In the first set of experiments in anesthetized non-pregnant ewes, we measured the spectra of several synthesized General MIDI music sounds created by a Korg X5 multitimbral synthesizer. Ewes were situated in a sound attenuated booth and a calibrated miniature hydrophone (Bruel and Kjaer, Model 8103) was placed in the abdominal cavity 5 cm from the wall proximal to the speaker (P). Analyses were performed with 1/3 octave bands from 65-10,000 Hz by a dual channel spectrum analyzer (Bruel and Kjaer, Model 2123).

In a second set of experiments, a professional musician produced predetermined notes on a trumpet and flugelhorn (F4,A4,C5,F5,A5,C6). Sound pressure levels were measured for the fundamental frequency and the overtones up through 3000 Hz. The amplitude and spectral signatures of the notes were recorded with intra- and extraabdominal hydrophones for off-line analysis. Finally, several bars of a popular tune were played on these two instruments. Identification of the instrument being played was attempted by playback of taped intrauterine recordings.

RESULTS
In the initial experiment with anesthetized ewes, attenuation at frequencies below 300Hz was low (Fig 1). Attenuation values averaged across several instruments showed nearly linear growth with increasing frequency. In some cases, musical notes below 300 Hz produced sound pressures which were greater inside the abdomen than they were outside the abdomen. Between 315 and 2500 Hz, the attenuation increased fairly consistently at a rate of 5 dB per octave.

The analysis of the sounds in air showed the expected greater contribution of higher frequency components for the trumpet. These components of the spectrum, mainly between 1500-3000 Hz, were reduced significantly when intrauterine recordings of trumpet sounds were analyzed (Fig.2). Somewhat smaller reductions in spectral energy in this frequency region were observed for the flugelhorn. The multibar patterns of trumpet tones appeared to the listener be more “flugelhorn-like”.
DISCUSSION

Filtering is a significant process that influences the quality, and thus the recognition of specific musical sounds. The result of a low-pass, high-frequency cut-off filter, characteristic of the abdominal tissues, is a generalized "muffling", noted previously in intrauterine recordings of voices (5,8). However, while there can be significant differences in externally-generated speech based upon the register of voice (6), the intelligibility of phonemes can be relatively high (5). Such a conclusion can be reached empirically by listening to taped in utero recordings of various musical passages (1). The present experiments show how this dampening is represented in the spectra of audible sounds generated by musical instruments.

The results reported above were based upon intraabdominal and intrauterine recordings of sound pressure levels. Sound transmission to the fetal inner ear is certain to be via bone conduction (4). The loss in energy is less than 10 dB for tones below 300-500 Hz, but may exceed 40 dB at higher frequencies within the audible range.

Thus, it is likely that the fetus will not be privy to overtones of most instruments at high frequencies. Rather, perception would be limited to melodic patterns and rhythms, and would be non-specific with regards to instrument identity.
REFERENCES


Spatial Performance as a Function of Early Music Exposure in Rats
(Rattus norvegicus)

Frances H. Rauscher, K. Desix Robinson, and Jason J. Jens
University of Wisconsin, Oshkosh

Introduction
Evidence is accruing from neuroscience laboratories all over the world that early enrichment can have a profound effect on cognition. Supporting this claim, research suggests that early musical experience may improve learning in other intellectual domains, particularly spatial domains (Rauscher, Shaw & Ky, 1993; 1995; Rideout & Taylor, 1997). For example, three-year-old children provided with six months of keyboard training scored significantly higher on items measuring spatial-temporal reasoning than control groups (Rauscher, Shaw, Levine, Wright, Dennis & Newcomb, 1997).

A neurological explanation for these findings can be found in the “trion” model of the cortex, developed by Gordon Shaw and his colleagues (e.g., Leng & Shaw, 1991). The trion model proposes that musical activity strengthens neural firing patterns that are also exploited by spatial-temporal tasks. Shaw suggested that music training provided to young children may facilitate learning in other spatial domains, such as mathematics and engineering. Consistent with Shaw’s model, EEG coherence data taken from subjects listening to a Mozart sonata and then performing a spatial-temporal task revealed a carry-over of coherence patterns from the music to the task in two brain regions (Sarnthein, vonStein, Rappelsberger, Petsche, Rauscher, & Shaw, in press), providing further support for a neurophysiological basis.

One approach to identifying the potential neural mechanism governing music’s effect on spatial-temporal reasoning is to establish its existence in an animal population. Despite a growth in research exploring the link between music and spatial reasoning in humans, studies regarding the effect of auditory enrichment on animal behavior are, to our knowledge, extremely rare. Although several researchers have explored the effects of enrichment on brain structure and function (see, for example, Rosenzweig, 1966) there is a dearth of research focused on auditory enrichment, particularly music. Magnetic resonance imaging (MRI) has shown that humans with either perfect pitch or early music training have larger left auditory cortices than humans who have not received formal training (Schlaug, Jancke, Huang, & Steinmetz, 1994; Schlaug, Jancke, Huang, & Steinmetz, 1995), suggesting the existence of experience-dependent neural plasticity in humans. Although studies have demonstrated that enrichment-induced neurogenesis in rodents can enhance learning (Rosenzweig, 1966, Kempermann, Kuhn, & Gage, 1997), none have examined the effects of musical enrichment.

Two areas of research have encouraged the current study: (1) research in humans suggests a neural representation for music-induced enhancement of spatial-temporal reasoning based on Shaw’s trion model, and (2) studies indicate that environmental enrichment can cause neurogenesis in the hippocampus of the rat. To search for a neural mechanism that controls spatial-temporal enhancement in humans, we conducted a study to determine if the enhancement in spatial performance—primarily a hippocampal function—found through music enrichment in humans could be found for an organism with a much simpler neural code: the Long-Evans rat (Rattus norvegicus). Positive findings would suggest new directions for research into the neurophysiological basis of the causal enhancement of spatial-temporal reasoning through music.

---

1 spatial-temporal reasoning, a subset of spatial reasoning, is the ability to perform mental operations on objects in space and time.
Procedure
Because it is evident that embryonic sensory stimulation can affect postnatal capabilities, and given findings indicating that rodents can discriminate different musical stimuli (D'Amato & Salmon, 1984; Poli & Previde, 1991), we randomly assigned 90 rats in-utero to three auditory exposure conditions: (1) the Allegro con spirito of Mozart's Sonata for Two Pianos in D Major, K. 448 (8 min 24 sec duration); (2) the beginning of Philip Glass' Music With Changing Parts (also 8 min 24 sec duration), and (3) white noise (sound which contains all frequency spectrums). The Mozart sonata, a complex composition containing several interwoven musical motives, was previously shown to produce short-term enhancement of spatial-temporal reasoning in humans when compared to the Glass composition (a particularly repetitive work), or silence (Rauscher, Shaw, & Ky, 1995). We chose white noise rather than silence for our control group to obscure the sounds of the animal colony workers.

All rats were bred from stock obtained from Harlan Sprague-Dawley (Indianapolis, IN), and were housed post-weaning in same-sex pairs in 28 cm X 21 cm X 19 cm stainless steel cages in separate rooms. Each musical selection was looped during the animals' dark (active) cycles for 12 hours of continuous exposure per day in utero (approximately three weeks), continuing sixty days after birth. Sound levels registered 65 decibels at the animals' cages.

At age 58 days, the rats were weighed and put on a 23-hour food deprivation schedule to increase motivation. At age 59 and 60 days, the animals were placed in the alley leading to the goal box of a half-size Stone 12-unit T-maze (Stone & Nyswander, 1927) to accustom them to the routine of maze running and food reward. A diagram of the maze is provided in Figure 1.

At age 61 days we tested the rats' spatial performance in the maze. During each trial, the rats were exposed to either the Mozart Sonata (n=30), the Glass composition (n=30), or white noise (n=30), all played at 65 decibels, yielding a three by three experimental design. The conditions are summarized in Table 1.

Table 1
Assignment of Rats to In-Utero and Maze Exposure Conditions

<table>
<thead>
<tr>
<th>Maze Exposure</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mozart</td>
<td>10</td>
</tr>
<tr>
<td>Glass</td>
<td>10</td>
</tr>
<tr>
<td>White Noise</td>
<td>10</td>
</tr>
</tbody>
</table>

Figure 1. Floor plan of the Stone multiple T-maze (half-size), showing doors and pseudo doors. The doors were closed behind the animal to prevent excessive retracing. The pseudo doors were included to make all pathways containing doors alike at each end, so that the animal could not choose a path on the basis of the presence or absence of a door.

At age 61 days we tested the rats' spatial performance in the maze. During each trial, the rats were exposed to either the Mozart Sonata (n=30), the Glass composition (n=30), or white noise (n=30), all played at 65 decibels, yielding a three by three experimental design. The conditions are summarized in Table 1.

Table 1
Assignment of Rats to In-Utero and Maze Exposure Conditions

<table>
<thead>
<tr>
<th>In-Utero Exposure</th>
<th>Mozart</th>
<th>Glass</th>
<th>White Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mozart</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Glass</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>White Noise</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

689
Each animal participated in a maximum of 15 trials or until criterion (three successive errorless runs) was reached. Testing was performed blind. A trial ended after the rat either reached a Froot Loop (sugar cereal) reward in the goal box, or after four minutes, whichever came first. Testing took place over a period of five days, three trials per day. Animals alternated in groups of three to provide a rest period for each rat between trials. All trials were videotaped.

Results
Dependent measures included total working time in the maze (WT) and total number of errors, including retracings (E&R). The animals were allowed to venture four inches into a blind alley before an error was counted against them. To ensure objectivity, working time in the maze was measured by a timer operated by two laser photoeyes attached to the wall of the maze two inches outside of the start box (timer start) and two inches inside of the goal box (timer stop). Errors were tabulated from the videotapes by a coder blind to all experimental conditions. Inter-rater reliability was calculated at $r = .99$.

An alpha level of .05 was used for all statistical tests. To compare group means, we performed a three (In-utero Exposure: Mozart, Glass, white noise) by three (Maze Exposure: Mozart, Glass, White Noise) two-factor between-subjects analysis of variance (ANOVA) on both WT and E&R. Main effects were found for both dependent variables for In-utero Exposure only ($F_{(2,81)} = 3.53, p = .03$ and $E_{(2,81)} = 10.42, p = .001$, respectively). No other main effects or interactions were found for either variable. The rats who received the music of Mozart in utero and 60 days after birth ran significantly faster ($M = 34.72$) and made significantly fewer errors ($M = 2.0$) than the rats who received Glass (WT $M = 50.11$, E&R $M = 2.85$) or white noise (WT $M = 44.29$, E&R $M = 3.35$). Unpaired t-tests found that the Mozart group differed from both the Glass and white noise groups ($t_{(1,58)} = -2.41, p = .02$ and $t_{(1,58)} = -2.13, p = .04$), which did not differ from each other ($t_{(1,58)} = -0.91, p = .37$). Exposure to the music/noise in the maze during running did not affect either variable. The WT and E&R means for In-utero exposure are graphed with their standard errors in Figure 2.

![Figure 2](image-url)  
**Figure 2.** WT & E&R means and standard errors for rats who received in-utero plus 60-days exposure to either Mozart, Glass or white noise. Animals in the Mozart group ran significantly faster and made fewer errors than animals in either the Glass or white noise groups. The Glass and white noise groups did not differ.
To determine if the differences we found for \textit{in-utero} exposure were due to learning versus something else (i.e., stress) we examined the data for both WT and E&R on a day by day basis. (Note: We ran three trials per day.) These data are graphed in Figure 3.

A two-factor ANOVA (with \textit{In-utero} Exposure as a between-subjects factor and Day as a within-subjects factor) performed on WT found significant main effects for both factors ($F_{(2,267)} = 9.07, p = .0002$ and $F_{(2,267)} = 24.48, p = .0001$, respectively). To determine if there were differences between groups on each day, we next computed individual one-factor (\textit{In-utero Exposure}) ANOVAs for each day, and discovered that the groups did not differ on the first day of training ($F_{(2,267)} = .41, p = .67$). On Day 3, however, the ANOVA was significant ($F_{(2,267)} = 3.04, p = .05$), with the Mozart group differing significantly from the Glass group ($t_{(2,178)} = 4.99, p = .01$) - a difference that increased in magnitude through Day 5 ($t_{(2,178)} = 6.38, p = .001$). The Mozart and white noise groups differed significantly on Days 4 and 5 only, whereas the Glass and white noise groups did not differ on any day. One-factor repeated measures ANOVAs performed on Day for each exposure group indicated that all groups showed significant learning during the course of the experiment.

The data for E&R showed similar trends, with the two-factor ANOVA also finding significant main effects ($F_{(2,267)} = 20.87, p = .0001$ and $F_{(2,267)} = 27.97, p = .0001$, respectively). However, unlike WT, the Mozart group differed from the white noise group from Day 1 ($t_{(2,178)} = 5.25, p = .01$), although it did not differ from the Glass group. As with WT, the Mozart group began to differ significantly from the Glass group on Day 3, while the white noise and Glass groups did not differ on any day. Again, one-factor repeated measures ANOVAs performed on Day for each exposure group indicated that all groups made significantly fewer errors as the study progressed.
Discussion
This research suggests that early musical enrichment can enhance spatial performance in the rat, and strongly supports the existence of a neurophysiological basis for the enhancement found in humans. Rats who were exposed to Mozart's Sonata for Two Pianos in D Major, K. 448 in utero and for 60 days after birth performed significantly better in a multiple T-maze than genetically identical strains of rats who were exposed either to Philip Glass' Music With Changing Parts or to white noise. The Mozart group ran faster and made fewer errors overall. Although all three groups of rats showed significant improvement in their ability to navigate the maze over the course of the study, the animals who were exposed to the Mozart learned faster, as indicated by both WT and E&R. That WT did not differ for the three groups at the outset of the experiment suggests that the effect was due to learning rather than to some artifact of the experiment.

These data, however, should be interpreted with caution. The difference between the Mozart and white noise groups in number of errors on Day 1 suggests that perhaps stress, rather than enrichment, contributed to the number of errors made. To explore this hypothesis, we will test a group of rats exposed to silence instead of white noise. These data will provide us with information regarding the value of the white noise animals as a control group. Given the strong similarities in the trends of both dependent variables, however, this alternative explanation seems unlikely.

Exposing the groups to the three types of auditory stimulation during testing had no effect on performance. This finding was not surprising; it seems unlikely that short-term auditory exposure would affect rats as it does humans. However, in drawing comparisons between this study and Rauscher et al.'s (1993; 1995) work with humans (which showed significant increases in students' spatial-temporal performance after listening to the Mozart sonata), it is important to acknowledge that the animals in the present study were exposed to the music during performance of the task, whereas the humans were exposed to the music prior to it. The duration of exposure varied in the two studies as well. Whereas Rauscher et al.'s students attended to the music for ten minutes, the rats in the present study were exposed to the music for a maximum of four minutes, usually less, depending upon how quickly they learned the maze.

These data compliment the recent findings of Kempermann, Kuhn & Gage (1997), who demonstrated that environmental enrichment can induce hippocampal plasticity and improve maze performance in rodents. Because both studies demonstrated improved spatial performance, it is tempting to conclude that the behavioral enhancement was induced by neurogenesis in the hippocampus. However, we must again recommend a more conservative approach. Further research is needed to determine if areas of the brain in addition to or other than the hippocampus were affected by musical enrichment, and if other learning tasks can be affected in addition to spatial tasks. We are currently in the process of performing the histology necessary to address the former question. It is also important to note, as pointed out by Kempermann et al. (1997), that the "enriched" conditions in laboratory studies such as these must still be considered deprived conditions compared to what the animals would experience in the wild. Therefore, questions of ecological validity will need to be addressed.

Further studies are also needed determine the musical components that are relevant to the effect for both humans and rats. Although there are many similarities between human and animal sound processing, there are also significant differences. For example, although rats can discriminate complex auditory stimuli, their discrimination appears to be based on timbre alone (Poli & Previde, 1991). Furthermore, they appear to be incapable of tracking pitch relationships, although they can utilize frequency cues for discrimination (D'Amato & Salmon, 1984). Humans, on the other hand, rely much more strongly on melodic contour for discrimination (Trehub, Bull & Thorpe, 1984). Regardless of these differences in perception, it is still possible that the neurophysiological mechanisms involved in the enhancement of spatial performance through music are similar.
Another possible area of investigation relates to dose. The animals in the present study were provided with musical enrichment for 12 hours/day in gestation and 60 days following birth. It would be interesting and worthwhile to determine if lesser doses of music would induce similar behavioral improvements. A study is currently underway to examine the effects of various lengths of musical exposure on hippocampal neural development. Behavioral effects will be tested in additional studies based on those findings.

The recent state of research and theory in comparative cognition is strongly rooted in the experimental investigation of precisely controlled and recorded animal behavior. Although it would be foolish to conclude on the basis of these data that there are no substantial differences in the effects of music on human and rat spatial cognition, it seems likely that further work in this area will produce a greater appreciation of the biological mechanisms of cognition that are common to both species. The findings offered here contribute to our understanding of the common mechanisms that underlie human and animal behavior by providing behavioral evidence that the spatial task improvement found for humans through early music exposure may be regulated by neurophysiological factors that also mediate musical processing and spatial performance in the lowly rat.

References


Music Performance: Motor Processes
The present paper reports on two investigations. Both concern the relationship between the motor patterns involved in music performance and the way musical structures are apprehended by the musician. The first is a survey intended to demonstrate the great differences in motor patterning, and by extension in mental representation as well, that are encountered when the "same" music is played on various instruments. The second is an in-depth study dealing with possible fingerings and their interpretational significance as they emerge in two short piano passages. The two investigations are complementary and bring the same message: knowledge of the motor aspects of music-making is an essential component in our attempts to form a comprehensive understanding of music, a component that has largely been neglected in music analysis and music aesthetics. There is also a differential psychology of musical performance that awaits closer, systematic study.

Passages of music usually exhibit a more or less orderly internal organization in terms of repetition, sequence, variation, complementation, balance, symmetry etc. When rendered on an instrument, however, this musical structure is broken against the muscular functions of our anatomic apparatus and the basic conditions of sound production associated with each instrument, as well as against the layout of the tones on the instrument and the technical solution chosen by the musician in each specific case. These conditions join to make up a motor or — to use a perceptual term — a proprioceptive structure that somehow coexists with the representation of the musical structure.

In order to gain an idea of the variety of motor patterns inherent even in simple musical structures, a number of professional musicians, playing instruments characterized by entirely different man-to-sound interfaces, were asked how they would play four conventional musical configurations: a two-octave up-and-down major scale, a two-octave up-and-down root position major triad, and finally rising/falling diatonic and chromatic sequences built on the scale and triad, respectively. Given C major notations of the configurations, the players were asked to describe their execution in detail using the signs of their trades. They were also incited to indicate alternative ways of execution and to transpose the examples to keys necessitating radically different technical solutions. In the ensuing interviews the musical properties of the various ways of playing as well as their proprioceptive qualities were discussed.

Since the main interest of the outcome is the extent to which the structural organization of the music matches the motor representation of it, we must first establish some basic properties of the stimulus configurations. The scale and the triad consist of a rising portion interlocking with a corresponding falling portion. Although rising and falling motions are far from the same thing perceptually, the pitch identity in these specific cases makes for a strong sense of reversal, indeed mirroring. Because the scale and triad
both start from root position, the pitch repetition at octave distance is quite
unmistakable, and therefore the ascent and descent appear to be subdivided into two
patterns issuing from tonic notes. Turning to less patent units, the rising scale may
seem to consist of two disjunct tetrachords. The configurations built on the one-octave
scale and triad, have several inherent sequential organizations, depending on whether
what we may call downbeat or upbeat grouping seems to dominate.

A pianist has to cope with the irregular white-and-black topography of the keyboard,
and the functional non-equality of the thumb and the other fingers, as well as with the
conflict between the left-to-right pitch lateralization of the keyboard and the mirror
thumb-opposite-thumb symmetry of the hands. Scanning the results pertaining to the
scale and the triad, it appears that right and left hand motor representations are never
identical, and that (considering the hands separately) the octave identity of the material
is always retained. But since the changes of hand position in the right and left hand
never coincide, this identity is nevertheless affected: depending on hand and direction of
pitch motion, the position units either begin or close with the tonic note. The mirroring
between rising and falling motion is obscured, because sub-positioning the thumb
involves a motion that is quite different from that of super-positioning the middle or
ring finger. The tetrachordic make-up of the scale is not reflected since the fingering
groups are of different length. Some transpositions involve radical changes in motor
patterning, affecting for instance the proprioceptive expression of the tonic note.
Turning finally to the sequences, fingerings that orderly reflect the melodic inflections of
the recurring motifs can sometimes be devised, sometimes not.

When playing the violin, the left hand and the right arm have entirely different
functions. The left hand intonation allows for many interdependent options with
respect to choice of string, hand position along the neck, and fingering, while the right
arm technique (besides all niceties associated with tone quality) offers a great variety of
bowing patterns. These are chosen to fit the tempo and meet the demands of the
musical context, and are therefore disregarded in this short account. When playing for
instance a rising scale, the many intonation options can be used to yield quite diverse
technical solutions — one may wish to retain a certain position, to keep to a certain
string, to use or avoid open strings, to find fingerings sequences that recur or reflect
melodic inflections. The long and heavy strings of the double-bass make for short
coherent units of fingering, frequently within quite high positions. The thumb is used on
occasion, and so are loose strings and flageolets. Guitar playing offers a further kind of
patterning, complementing the motions of the left hand: four fingers of the right hand
are used to form convenient plucking patterns. Due to the tunings of these three
instruments, octave equivalence has no proprioceptive counterpart, and transpositions
tend to be associated with entirely new left hand patterns. Rising scales and triads are
proprceptively different from falling ones, because the changes of position take place
at different notes.

The holes and keys of woodwind instruments like the flute or the clarinet are operated
according to complex systems of combinations, and a shift from one note to another
may correspond to simple or to quite complex finger motions: sometimes just one finger
is lifted or depressed, sometimes several fingers of both hands are involved, closing
some holes and opening others. Due to this fact, the possibilities to devise sequences of
similar motor patterns corresponding to recurring musical motifs are slight, and for the
same reason transpositions are bound to imply great proprioceptive differences. When overblowing to reach higher registers, the same fingering combinations recur, and since the flute overblows in the octave, flute-playing is (to some extent) characterized by motor similarities at octave distance. Rising scales are (especially on the flute) different from falling ones, because the former generally involve lifting the fingers, while the latter is associated with depressing them.

Leaving again the concurrent and intricate motions of lips and tongue out of account, the notes to be produced on brass instruments are selected by overblowing and by lengthening the tube, but the valve adjustments of the trumpet, engaging three fingers, are very different indeed from the arm/slide positions of the trombone. The overlap of the higher overblowing registers, allowing the same note to be played with different valve combinations and slide positions, respectively, is an important technical resource. Generally speaking, tones at octave distance may exhibit identity (as far as finger combination or arm position is concerned) depending on how the player chooses to play them. The way down scales and triads is different from the way up due to various locations of the overblowing shifts. Transposition means considerable changes of motor pattern — some scales and especially some triads can be played with little finger or arm involvement, while others demand frequent adjustments. The prospects for reflecting sequenced motifs with a recurrent, similar motor pattern are fairly good, particularly on the trombone.

The organ pedal, finally, exemplifies an altogether different kind of playing conditions. Using tips or heels, the keyboard is basically played with left-right alternation, to which is added shifts from tip to heel of the same foot, shifts of foot on the same key, as well as sliding movements from key to key. Apart from considerations of convenience, the main factors when devising the playing patterns are tempo and articulation. If strict legato is to be maintained, the motions become more complex — such patterns are somewhat slow but do suggest a feeling of legato. Descending motions tend to be different from ascending ones, and while octave transpositions may be played as proprioceptively identical, other transpositions generally bring new motor patterns.

The general conclusion is that transpositions (including in many cases even octave transpositions), inverted motions, and reiterations of identical or closely similar motifs have few and imperfect correspondences in the proprioceptive domain. The bodies of musicians deal with other structures than their ears and eyes, and to the extent that all these representations of the music coexist in their minds, they certainly have a rich and bewildering experience of it.

Two short Chopin excerpts will serve to illustrate the second investigation to be reported.

The Prelude in B major Op. 28 No. 11 starts with a right hand melody, whose normal fingering is easy to find: 3|234314|3. But it seems preferable to use a strange and (at first) quite difficult fingering: 3|235454|3. Why choose a fingering that defies the anatomy of the hand? The following bars can be conceived of as a two-voice structure,
as a kind of swift dialogue between D# and C#, played by the little and the ring finger, and a murmuring lower strand insisting on F# and G#, left to the other fingers. The preferred fingering prepares for this polyphonic separation to come. But there is a further reason. In order to master the sub-positioning of the little finger, a certain lateral adjustment of the hand/arm is necessary. This makes both for a relaxing suppleness in the wrist and for some favourable tension in the fingers, and the latter fact bears on the interpretation of the passage. The focussing on the precarious finger work seems to give the melody a peculiar, somewhat pressed intensity — at least to the one who plays, it sounds quasi oboe.

In order to understand the point of the second example, we have to make some preliminary analytic observations. The posthumous F-minor etude starts with a twice repeated, open-ended six-note motif, mainly made up of half steps. (There is an additional, latent minor second between the first and the fourth note, which brings a resolution to the dissonance between the first and the third note.) A B♭ minor triad may be heard, but it yields immediately to a transient dominant harmony when the fifth is resolved to an augmented fourth. Rhythmically the anacrustic six-note motif may be divided into two three-note sub-motifs: first a compressed chromatic rise, then a more expansive gesture that from a phenomenological point of hearing suggests either a sense of frustrated attempt at breaking out, or a sense of a rounded shape enclosing the first sub-motif. After the repeat, the music seems to offer us a third statement of the six-note-motif, but this time the breaking out/enclosing gesture is replaced by the first sub-motif, which is allowed to expand and form a rising sequence that eventually brings the introductory melody to a close. It thus appears that the germ of development was hidden in the first compressed sub-motif, and that the sense of breaking out, inherent in the second gesture, is therefore less preferable. It is also clear that the first transformation C—D—F discloses what the starting motif itself enclosed: the tonic note C.

Now, which fingerings are possible for this melody, which proprioceptive meanings are inherent in them, and which phenomenal musical structure do they support? Disregarding possible fingerings expressing the non-preferred aspect of attempt-at-breaking out, we have two fingerings to consider: 123 254 and 213 254. Both embody a proprioceptive sensation that corresponds to the preferred sense of content-being-enclosed — a pianist playing in this manner begins with a very contracted hand position, which is then covered by a wider position when playing the B♭—F skip with the index and the little finger. But what about the first three notes? The straight fingering 123 may seem better, since it makes for a 1—2 finger progression reflecting the latent B—B♭ motion. On the other hand, the combination 213 posits the thumb under the hand already during the first sub-motif, where it will stay for the rest of the larger six-note unit. This fingering tends to inhibit the rising aspect of the initial sub-motif, and makes the enclosed core, the prospective dominant tone C, well-defined in hand as well as in thought.

Turning to the release of the melody in m. 3, there are three fingering options, corresponding to three distinctive phenomenal aspects and to three distinct interpretative options. The rising sub-motif can be played with the straight fingering 123 all three times, promising a rising continuation and turning the intervening second sub-motif into a diversion or blockage. If this alternative is chosen, the pianist has to
infuse the repeats of the initial motif with a gradually increasing impetus, that explains why the hand and the music eventually comes up with the expansive mutation C–D–F. Conversely, if you play the 213 fingering all three times, you have chosen a mode of playing that has the inherent tendency to bring the motion of the B–C–D♭ particle back to C, and that postpones all initiative until the transformed motif, played of course with a straight fingering. Finally, having twice started 213, the third occurrence of the crucial sub-motif may be played 123. This option is associated with a strong sense of interference in the melodic process: the release is located to a stage prior to any sign of actual change in the melody. Proprioceptively this means that the motion, that until now has imprinted the sub-motif and only resulted in circularity around C, has been abandoned, and that a "new", decisive motif has emerged, starting with the thumb from B, the leading note of C.

We have seen that various phenomenal musical configurations correspond to certain fingerings, and that, far from being just a matter of manual convenience, fingering patterns, due to the proprioceptive meanings they suggest, support and give rise to specific interpretations, to differences that may readily be heard. It should finally be pointed out that such subtle relationships between bodily sensation and musical content are highly vulnerable to transposition. Little of what has been said above about this etude is true if the introduction is raised by a semitone: the "corresponding" F♯ minor etude is another work, and probably a work of inferior value.
Interdependence of right and left hands in sight-read, written, and rehearsed fingerings of piano music in parallel octaves

Richard Parncutt, Keele University
John A. Sloboda, Keele University
Eric. F. Clarke, Sheffield University

Introduction

Passages in which the left and right hands (LH and RH) play in parallel at a distance of an octave occur commonly in piano music. They pose a specific technical problem to pianists: because the hands are symmetrical with respect to each other, quite different fingerings are required for the two hands. These two fingerings must be determined not only simultaneously, but also several notes in advance of execution, to enable forward planning of changes of hand position.

In sight-reading such passages, it would clearly be impossible for a pianist to work out fingerings for the two hands from first principles (e.g., the fingering "rules" of Parncutt, Sloboda, Clarke, Raekallio, & Desain, in press). Rather, it appears that pianists spontaneously access a sizeable "cognitive library" of fingering patterns that have been acquired gradually over years of pianistic training. Sources include technical exercises (including scales and arpeggios), fingerings printed in scores, principles of fingering taught by teachers and read about in books, and fingerings worked out creatively by individual pianists for pieces in their repertoire (Clarke, Parncutt, Sloboda, & Raekallio, 1997).

In the present study, six pianists performed two pieces of piano music in which LH and RH play in parallel octaves throughout and for which no widely agreed upon fingerings exist: two of Czerny's Kurze Uebungen Op. 821 (No. 54 in C#, No. 66 in g). Pianists first performed at sight, and later, on a separate occasion, after rehearsal. Within each condition they played each piece with RH alone and then with hands together. At the second session, they also supplied their preferred fingerings for each piece in written form.

On the basis of our intuitions as pianists, together with results of structured interviews conducted with professional pianists (Clarke et al., 1997) and comments from the pianists participating in this experiment, we hypothesised in advance of data analysis that:

1. In all conditions, RH fingerings would be ergonomically easier than LH fingerings. Rationale: It is difficult to attend to the fingerings of both hands at once. In parallel melodic motion, pianists tend to focus on RH and allow LH to follow, especially in sight-reading. This behaviour was particularly encouraged in the present experiments in which pianists first played RH separately but not LH separately.

2. For both hands, written fingerings would be ergonomically easier than rehearsed, and rehearsed easier than sight-read. Rationale: It takes time to arrive at a good fingering when there are many different possible fingerings to choose between. In recent work we algorithmically generated hundreds of "playable" fingerings for melodic fragments taken from these pieces (Parncutt et al., in press).

3. Changes of hand position would more often be synchronous in sight-read than in rehearsed fingerings, and in rehearsed than written fingerings. Rationale: Written fingerings may be ergonomically optimal for LH and RH separately, without allowing for synchronous
position changes. Rehearsal enables a greater number of asynchronous changes than sight-reading, allowing the fingerings in each individual hand to be ergonomically easier.

4. Rehearsed RH and LH fingerings would be more likely to deviate from written fingerings when written fingerings involved asynchronous position changes, because deviations tend to line up hand position changes. Moreover, on the basis of findings of Sloboda, Parnicut, Clarke, and Raekallio (1997), we expected that the number of note errors would increase as the number of departures from written fingerings increased.

5. There would be less disruption of RH by LH in rehearsed than in sight-read performances (measured by the difference between RH fingerings for RH alone, and RH fingerings for hands together).

6. The rate of errors would be greater near asynchronous changes than near synchronous changes. This tendency would be more pronounced in sight-reading than in rehearsed performance.

Method

Participants. Six pianists were recruited by mail and were paid for participation. All had university-level diplomas in piano performance. On average, they had played (practised and performed) regularly for 27 years (sd = 13 years; 3 pianists had played for 30-50 years, and 3 for 15-17 years). Two of the pianists were giving regular public performances (50-60 per year); the others performed relatively seldom (0 - 5 public performances per year).

Materials. Each pianist played two pieces from Czemy's Kurze Uebungen Op. 821: No. 54 in C# major, and No. 66 in G minor (identified as "Piece D" and "Piece F" by Parnicut et al., in press). Each piece involved hands together playing fast running melodic 1/16-notes, an octave apart, for 7.5 measures of 4/4 metre; in all, there were 121 notes in each hand. The two pieces are dominated by arpeggio-like figurations. The Study in C# major begins with a rising scale in broken thirds: in the RH, the notes are C#4 E#4 D#4 F#4 E#4 G#4 F#4 A#4 (etc.); an octave lower in the LH. The Study in G minor begins with a jagged rising arpeggio: in the RH, the notes are G3 Bb3 D4 Bb3 G4 D4 Bb4 G4 (etc.); again, an octave lower in the LH.

Procedure. The pianists performed on two separate occasions. In July 1995, they were presented the pieces for the first time and asked to sight-read them. In April 1996, they were sent scores of the pieces by mail in advance of their performances. On both occasions they first played the Study in C# major with RH alone, then with hands together, then the Study in G minor RH alone, then with hands together. On both occasions they performed one or more different tasks before repeating all four performances.

Raw data. Performances were recorded as MIDI data and as birds-eye videos of the player's hands. Fingerings were transcribed by watching the video in slow motion. The transcriber also recorded errors (by ear). Additional data included written fingerings.

Design. The 6 pianists each played the 2 pieces in 2 different "handings" (RH alone then hands together), in 2 repetitions, under 2 different conditions (sight-read and rehearsed). In all, there were 96 performances. For each combination of piece and condition, only the repetition with fewer note errors was fingered and analysed. If the same number of note errors occurred in both repetitions, the second version was used. Our analysis is based on these 48 performances.
Results

Tempo. Performers were free to choose an appropriate tempo, which was then determined
from the total duration of each performance, from the onset of the first note to the onset of
the last. Considerable tempo variation was observed among the six pianists, $F(5, 42) = 22,$
$p < 0.0001$. As expected, rehearsed performances were played faster than sight-read, $F(1,$
$24) = 11, p < 0.005$; mean 1/16-note durations were 230 ms (sight-read) and 180 ms
(rehearsed). No significant difference was observed between the tempos of the two pieces;
nor did the tempo of performances with RH alone did not differ significantly from
performances with hands together. There was no interaction between piece and condition,
implying that the two pieces were equally difficult to learn; incidentally, nor was there any
significant difference between the number of errors made in each piece, implying that the
absolute difficulty of the pieces was about the same.

Playing intended fingerings. Fingerings differed across the three conditions (sight-read,
written, rehearsed). Differences were expressed by counting the number fingering
deviations; the hypothetical maximum score was 121, the number of notes in each piece. As
expected, the smallest difference between any 2 of the 3 conditions occurred between written
fingerings and fingerings in rehearsed performance, $F(2, 102) = 26, p < 0.001$. This
difference averaged 15 notes, by comparison with 33 notes for sight-read vs rehearsed, and
36 notes for sight-read vs written. The extent of fingering variations across the three
conditions did not vary significantly with the 2 pieces or with 3 “handings” (RH alone, RH
in hands together, LH in hands together).

Interference between hands. When comparing only written fingerings with fingerings in
rehearsed performance, however, differences did vary with handing, $F(2, 32) = 3.79, p <
0.05$. In the RH alone performances, the difference averaged 10.5 notes; and in the hands­
together performances, the difference for the RH was 12.7 notes, and for the LH 20.9 notes.
Thus, in the hands-together performances, the LH did not disrupt intended RH fingerings
any more than they were already disrupted by the act of performing, but in hands-together
performances, intended LH fingerings were disrupted more than intended RH fingerings,
consistent with Hypothesis 5 (above). With the present data we were unable to test the
hypothesis that deviations between written and rehearsed fingerings involve reverting to
habitual fingerings (Clarke et al., 1997).

Calculated difficulty. The difficulty of each chosen fingering (sight-read, written,
rehearsed) was calculated according to the model of Pamcutt et al. (in press). The model
predicts fingering difficulty as a sum of several components, of which only two were
considered here: one involving finger-spans that exceed their maximum comfortable size
(called the Stretch Rule in Pamcutt et al., in press), and the other depending on the number
of changes of hand position (the Position-Change-Count Rule).

The difficulty score accumulated by the Stretch Rule varied with piece, $F(1, 67) = 33, p <$
0.001, performance condition, $F(2, 67) = 6, p < 0.01$, and hand, $F(1, 67) = 22, p < 0.001$.
Regarding performance condition, the greatest average stretch score occurred in the sight­
reading condition (mean 52) and the smallest in rehearsed performance (mean 39), - but not,
as we expected, in written fingerings (47). Hypothesis 2 was thus only partly confirmed.
This finding suggests that rehearsal allows pianists gradually to reduce the number of
unnecessary stretches, even beyond the number required to execute their own written
fingerings. Regarding the difference in calculated difficulty found between the two hands,
stretch scores for the RH (mean 54) exceeded scores for the LH (38), suggesting that
pianists were more likely to risk large stretches in the RH. We take this as evidence that the
pianists were focusing their attention on the RH, probably looking at it more than at the LH, and allowing the LH to follow the RH's lead. This finding contradicts, and sheds new light on, hypothesis 1.

One would expect that, if pianists were stretching further in the RH than in the LH, they would change hand position less frequently; but this prediction was not confirmed. The number of changes of hand position (counting 1 for a half change and 2 for a full change, as defined in Parn cott et al., in press; here, most changes were half changes) varied only with piece, $F(1, 67) = 20, p < .001$. It did not vary with hand, nor did it vary with condition.

**Synchronicity of changes of hand position.** The mean number of synchronous hand-position changes over all performances was 28, compared with 49 asynchronous changes. Of the asynchronous hand-position changes, a greater number (t test) occurred in the RH (mean 26) than in the LH (mean 23), consistent with the idea that LH followed RH (or more specifically, that LH position changes were more often determined by RH changes than vice-versa). In line with hypothesis 3, there was a small but insignificant ($p < 0.1$) decrease in the number of synchronous changes between sight-read and rehearsed performances, but clearly no difference between rehearsed and written fingerings. Hypothesis 4 was not confirmed: no significant correlations were observed between the number of asynchronous changes, the number of errors, and the number of deviations between written and rehearsed fingerings for each performance. Hypothesis 6 regarding the precise location of errors relative to position changes could not be tested statistically due to the relatively small numbers of errors.

**Conclusion**

Our main findings were that, in piano music where the hands move in parallel octaves, (i) intended LH fingerings were disrupted by hands-together performance more than intended RH fingerings, suggesting that pianists focus more attention on RH fingering; and (ii) RH fingerings involve more or bigger stretches between successive fingers than LH fingerings. The focus on the RH may be a consequence of two facts: first, at some level it is impossible to divide or evenly share attention between the two hands, and second, errors in the RH tend to be more perceptually salient, i.e. more noticeable to listeners, than errors in the LH.

**References**


705
The researchers in this study are interested in the way student and professional trombonists manipulate the trombone slide. The most frequently asked question of trombone players by non-musicians is, 'How do you know where to put the slide?'. To answer that question the researchers examined the literature available in selected trombone method books. Then, to get a more scientific answer they compared that information with data collected from student and professional trombonists using biomechanical techniques.

Editors and compilers of method books take one of three approaches to the subject of slide placement. The first is simply to state that there are seven positions and to list the notes that can be played in each position. The second approach is to have an illustration of the slide and to describe the recommended distances between each of the positions on the slide. The third is to say that the ear tells where the positions are on the slide. A publication using the first approach listed above is, *The Imperial Method for Trombone* by R.N. Davis (1962). On page seven a trombone slide is shown with the notes possible in each of the positions. However, there is also the admonition that, "The ear is the only reliable guide to the location of the true position for any note."

Three earlier publications provide recommendations about the distance the slide should travel (second approach). *Arban's Famous Method for Slide and Valve Trombone and Baritone* by Charles L. Randall and Simone Mantia (1936) has a picture of the trombone slide on page thirteen with the notes possible in each position and a table of recommended distances. *The Modern Trombonist* by Tommy Dorsey (1944) also has a chart listing, "the correct distances between the positions". Finally, *Method for Trombone* by V. Comette (1937) has the usual slide pictured on page thirteen with the explanation, "There are seven positions, advancing in half notes separated from one another about 3 1/2 inches" (8.89 cm). The predicted distances for each of these methods are shown in Table 1.

| Recommended Distances Between Slide Positions |
|-----------------|---|---|---|
| Slide Motion   | Arban | Dorsey | Comette |
| 1 to 2         | 8.3  | 7.9   | 8.9    |
| 1 to 3         | 16.5 | 16.5  | 17.8   |
| 1 to 4         | 25.4 | 25.4  | 26.7   |
| 1 to 5         | 35.6 | 34.9  | 35.6   |
| 1 to 6         | 47.0 | 45.1  | 44.5   |
| 1 to 7         | 58.4 | 55.9  | 53.3   |

The use of the ear is emphasized by more contemporary writers (third approach) such as Donald Knaub (1964) and Reginald H. Fink (1970). Knaub (p. 6) after writing that the slide
positions are approximately three and one half inches (8.89 cm) apart states, "These positions are not fixed and vary with individual instruments and the key of the composition. The ear is the final judge as to exact slide placement". Fink (p. 58) speaks of the ‘precise placement’ of the slide and "The exact placement of the slide is determined by the ear, but you can use your eye and the feel of the arm to help you."

In summary, examination of the pedagogical material on trombone performance makes it clear that there is not agreement on where the slide should be placed. This may not be surprising since experienced performers can alter pitch without changing slide position. This allows them to play correctly even when this slide is out of position. Nonetheless, it still remains to be seen how far performers really do move the slide. Our approach has been to collect normative information which allows us to look for similarities across trombonists and to compare expert performers with student performers.

Several studies demonstrate the utility of using biomechanical measures to assess trombone performance. Lammers (1983) compared seven professional trombonists and seven college students studying the trombone as they played a short etude which required subjects to move between the first slide position and each of the other six slide positions. Surface electromyography was used to observe activity in triceps and biceps of the upper arm and the extensor and flexor carpi radials of the lower arm. Electrogoniometers were attached to the elbow and wrist. Professional trombone players used less elbow angle and less muscle activity while moving the slide more quickly from position to position. In a second study, Lammers and Kruger (1991) used a triaxial goniometer to observe abduction-adduction and flexion-hyperextension of the wrist. Ten college level performers participated in this study. Subjects performed the same etude as in the earlier study. All subjects showed changes in both flexion-hyperextension and abduction-adduction of the wrist during performance. The more accomplished performers used the wrist more rather than less. This was most notable in the longest movements to the sixth and seventh positions. More recently, Kruger, Lammers, Stoner, Allyn, & Fuller (1996) examined differences in the velocity of slide motion in trombone players as a function of expertise. This study found that all players move the slide as quickly as needed to reach a position rather than moving it as fast as possible all of the time. Even so, professionals move the slide faster than students.

The present study uses descriptive biomechanical techniques to measure the distances trombonists moved their slide during performance. It was hypothesized that professionals would move more accurately than student performers and that this would be most evident when performers were asked to play faster tempos.
Procedure

Nine full-time professional and eleven college student trombonists were studied to discover the positioning and movement of the trombone slide. Performances on each of the musical exercises were evaluated for musical quality and number of audible errors by an undergraduate music student with performance experience who was blind to the background of each performer. Professional performers, not surprisingly, made fewer audible errors summed across all of the exercises and they performed more musically than either adult amateurs or college student performers. A Vernier Ultrasonic Motion Detector (sonic ranger) interfaced to a Hewlett Packard QS/165 microcomputer was used to measure distances on the trombone slide. A 30.5 cm circular target was attached to the end of each player’s trombone slide while the motion detector was placed in front of the subject to read the motion of the slide. A Peak-5 system which uses high speed videography to assess motion was used to verify data from the sonic ranger.

The subjects were asked to play two exercises, each at three different tempos. The first exercise was the B flat scale played both ascending and descending without repeating the top note. The first trial was in common time at 60 beats per minute and each note a half note. The second trial doubled the first speed (quarter note per beat) and the third trial doubled the speed of the second trial. The researchers were interested in discovering differences that might occur according to speed. Special emphasis for this part of the study was attended slide movements from sixth to fourth positions, fourth to sixth positions, fourth to second positions and second to fourth positions. These movements were selected because they represent motions that do not involve first position and should allow for the most variation among and between subjects in length of slide motions. The second study exercise was constructed to make use of the entire slide. Starting in first-position "F" the exercise consisted of two notes played on each note descending chromatically to each position from one through seven. The subject was asked to perform this exercise with the quarter note set at ninety-two beats per minute and then twice that fast (half note at 92 beats per minute). Finally, the subject was asked to play this exercise with the whole note set at 92 beats per minute.

Results

For each motion of the slide, distances were computed by finding the change in position from the beginning to the end of each motion. The mean differences between first position and each of the other slide positions for professional performers and for student performers are presented in Table 2 (found on the next page). These differences were computed from data collected while performers played the second study exercise with the quarter note set at ninety-two beats per minute. A 2 (level of expertise) by 6 (slide motion) ANOVA was used to test whether the distance traveled to each position differed as a function of expertise. There was a significant interaction between level of expertise and slide motion, $F(5,75)=4.56, p<.001$. As can be seen in Table 2, professionals differed from students mainly in longer motions – first position to sixth position and first position to seventh position.
Table 2

Observed Distances in Centimeters

| Slide Motion | Professionals | | | Students | | |
|--------------|---------------|---|---|---------|---|
|              | Mean n Standard Deviation | | Mean n Standard Deviation | |
| 1 to 2       | 9.7 9 2.4 | | 10.5 10 1.9 | |
| 1 to 3       | 19.3 9 5.3 | | 18.9 10 1.7 | |
| 1 to 4       | 25.9 9 4.8 | | 28.5 10 1.5 | |
| 1 to 5       | 36.3 9 6.7 | | 39.1 10 1.5 | |
| 1 to 6       | 47.5 9 3.0 | | 45.9 9 3.0 | |
| 1 to 7       | 58.8 9 4.1 | | 52.0 9 2.7 | |

To further examine the accuracy of slide placement, distances between slide positions when the subject was not starting at the first position were compared. Distances were again computed by finding the difference between the beginning and end of each motion. The findings reported below are based upon measures taken while subjects performed the first study exercise (a B flat scale) at three tempos. A 3 (tempo) by 2 (direction of motion) by 2 (level of expertise) ANOVA with two repeated measures was used to compare the distances performers moved the slide between the second and fourth position as well as between fourth and second position. Professional performers move the slide smaller distances (M=19.5 cm) between second and fourth position than student performers (M=20.4 cm), F(1, 12)=6.22, p < .05. The tempo of the exercise influenced the distance traveled, F(2, 26)=2.89, p < .08. The mean distances for the three tempos were 19.9, 19.8, and 20.4 cm respectively from slow to fast. Professionals did not differ from students in the distance they moved the slide as function of the direction of slide motion or as a function of tempo. A second ANOVA was used to examine movement between fourth and sixth position and sixth and fourth position. Professional performers again differed from student performers in the typical distance that they moved the slide, F(1,12)=5.33, p<.05. However, professional performers moved further between fourth and sixth position (M=20.4 cm) than student performers (M=17.8 cm). Tempo of also influenced the distance the slide was moved, F(2,24)=3.29, p<.06. The means for the slow, moderate, and fast tempos were 19.4, 19.3, and 18.0 cm respectively. There were no differences between professional and student performers as a function of tempo. Finally, performers were found to move the slide shorter distances which contracting the slide (M=18.2 cm) than when extending it (M=19.7 cm), F (1, 12)=25.63, p < .05.

Conclusions

Professional and student performers were found to differ in their choice of slide position. Differences were most notable in the longest positions where students appear to use the positions recommended by Cornette (which uses equal distance intervals between each position). Professionals choose positions more similar to those recommended in the Arban method.
However, it should be noted that both professionals and students move the slide further between first and second position than is recommended by any of the method texts.

The hypothesis that differences between students and professionals would be most apparent when exercises were performed at faster tempos was not supported. Tempo did influence motions between second and fourth positions and motions between fourth and sixth positions. Increased tempo caused performers to increase the distance between second and fourth position, but to decrease the distance between fourth and sixth position. The observation that distances vary as a function of tempo suggests a need to collect normative information about distances at several tempos, and most likely with different types of music.

Observation of differences between professional and student performers is consistent with our earlier research (Kruger et. al., 1996; Lammers, 1983; Lammers and Kruger, 1991). The finding that professionals move further than students at the far positions is consistent with the finding that they are also moving the slide at higher velocities. Knowing that professionals are moving further not just faster may tell trombone teachers to emphasize position at least as much as slide speed. Our observations indicate that professionals facilitate reaching the far positions by playing to one side, thus decreasing the distance they need to reach to get to positions six and seven.

Our professionals clearly played more musically, even on the simple exercises used in this study. Using biomechanical techniques to study musical performance is a useful way to look for factors which differentiate the highly successful performer. Further, research is needed to determine how trombone players place the slide – by using visual cues to slide distance, by using kinesthetic feedback, or by using their ears to sense pitch and tone quality.

References

Music Performance: Modelling
Multidimensional Dynamic Shaping
Jörg Langner, Musikhochschule Hannover

Introduction
The dynamic shaping of music has not been, up to this time, at the center of the interests of performance research. This applies to the pioneer works of the thirties of Seashore, as well as to the modern studies since the seventies for example of Gabrielsson, Clarke, Palmer and Repp. A reason for this may be the fact that the importance of the dynamics for a good performance is seen as secondary in relation to that of timing (e.g. Bengtsson, 1975), a view that is also found under musicians, when they refer to, for example, organists who have the ability to play lively and expressively on their instruments, although nearly exclusively the means of timing is available to them. In going against this argument we can actually say that only a small outermost percentage of all music that is made or heard daily throughout the world originates from such instruments. The flexible shaping of dynamics is obviously held in very high regard by musicians and listeners, and the question for the properties of "good" dynamic shaping gets its relevance from this fact. No simple answer to this question exists. Some attempts based on the view of a simple linking between dynamics and timing (e.g. Todd, 1992) seems to be problematical.

In this position, the need for qualified analytical methods appear. For a good analysis, we may not confine ourselves to the loudness relations of directly neighboring tones, but have also to deal with the extensive relationships. We find an important approach to the multidimensional analysis of loudness curves in Todd (1994), whose Gaussian filter-banks with a broad spectrum of time-constants form a model for the explanation of rhythmic grouping. The dynamics of music are obviously complexly organized in several levels or dimensions. A first, synthetic example may serve as an explanation and makes clear the demands which we must set for an analytical procedure. Figure 1 shows the loudness curve of a simple, isochrone series of drum beats.

![Loudness Curve of a Synthetic Example](image)

Fig. 1. Loudness curve of a synthetic series of isochrone drum beats.

Listening to this series of tones, what do we hear? It is quite clear: A two-level superposition of dynamic arcs of crescendo and decrescendo, one narrow arc with several repetitions and a large one extending over the whole series. We expect that a qualified analytical procedure is able to condense, so to speak, these two dynamic arcs from the given loudness curve.

Method
The loudness curve of a recording or a live performance always serves as input for the calculation. For the way in which to get these curves and for other detailed information about the whole procedure see Langner (in preparation).
The first central step of the calculation consists of several smoothings of the loudness curve with different degrees. The degree of smoothing corresponds to the so-called "window size" which fixes the diameter within which the values are used for calculating the mean. The effect of such smoothings shall be demonstrated with the help of figure 2. The initial point is a loudness curve similar to that of figure 1 (the gaps between the onsets are filled), and we can see some examples for smoothing with different window sizes. (The unit for the window sizes are seconds because the loudness curve represents a function in time.) We notice that one of these smoothings lead to a welcome result: With a window size of 1.00 seconds one gets a curve still consisting only of the one large arc of loudness. We conclude from this: Smoothing a loudness curve with a broad and dense spectrum of numerous window sizes should lead to detection of all wide dynamic arcs in the music.

The Effect of Smoothing

![Diagram](image)

Fig. 2. The effect of smoothing. The original curve is similar to the loudness curve of figure 1 (merely the gaps between the onsets are filled). Smoothing with a window size of 1.00 seconds leads to a curve showing purely the one wide arc of the original. Smoothing with a window size greater than the duration of the example results in a horizontal line representing the mean value of the whole curve.

The second central step is to calculate the slope in each of the smoothed curves at each point in time and to multiply this value with a factor depending on the corresponding window size. This procedure leads to the effect that those curves smoothed with small window sizes reflect only the small dynamic arcs. (It is not possible to explain it carefully within the bounds of this paper.)

By the means of both steps we reach our goal: To separate the dynamic arcs from each other. The result can be seen in figure 3, a so-called "dynagram". Each row belongs to a certain degree of smoothing, the range of window sizes extends here from 0.125 to 32 seconds, logarithmically spaced. Note that dynagrams do not depict loudness but changes in loudness (the slopes of the curves had been calculated!). These changes are represented by the means of
black shading. The dark shades correspond to crescendo, the light to decrescendo and the middle grey represents "constant loudness".

Dynagram of a Synthetic Example

![Dynagram of a Synthetic Example](image)

Fig. 3. Dynagram belonging to the loudness curve in figure 1 (dark shading: crescendo; middle gray: constant loudness; light shading: decrescendo).

As we had wanted, the two levels of dynamic shaping are clearly visible: We can find the one large dynamic arc in the area of middle window sizes, and we can see the small repeated arcs above. Note that we have visualized a striking phenomenon: It is possible to play and to percept a crescendo and a decrescendo at different levels simultaneously. Figure 4 is nearly identical with figure 3, but the horizontal lines are omitted. This version of dynagrams is preferred for the following examples because it permits us to detect more details in the graph.

![Dynagram nearby identical to that in figure 3, but without the horizontal lines.](image)

Fig. 4. Dynagram nearby identical to that in figure 3, but without the horizontal lines.

Results

Now we change to the analyses of real music. Figures 6 and 7 show the loudness curves, figures 8 and 9 the corresponding dynagrams of two different performances of the following rhythm:
Fig. 5. Score of the performed rhythm.

Fig. 6 and 7. Loudness curve of a professional and a non-professional performance of the rhythm notated in figure 5.

Dynagram of a Professional Drum Performance

Dynagram of a Non-Professional Drum Performance

Fig. 8 and 9. Dynagrams of a professional and a non-professional drum performance of the rhythm notated in figure 5 (dark shading: crescendo; middle gray: constant loudness; light shading: decrescendo). We can see the traces of the stronger and more extensive dynamic shaping of the professional. Note that the horizontal axis in these figures show no markings for the time unit (seconds) but for the onsets of the notes (compare with the score of the rhythm).
Figures 6 and 8 belong to a professional performance which is rated by subjects as "very good". The non-professional performance related to figures 7 and 9 is evaluated as "poorly played". The differences in the dynamic shaping came to light in the dynamgrams: The professional drummer uses generally the stronger dynamic changes, and his performance shows a wider shaping. In the area of the window sizes from 1 to 2 seconds, a clear structure of crescendo-decrescendo-crescendo is to be seen. The non-professional performance lets such a shaping be missed, the most intense shading is found at the very small window sizes which means loudness differences exist only between direct neighboring notes. The amateur "thinks" merely from tone to tone.

Figure 10 shows the dynamgram of a professional piano performance of Erik Satie's "Gymnopedie No.1". We see here the traces of a very wide dynamic shaping: Deviations from middle gray are found down to a window size of 64 seconds. Moreover, the formal structure (two identical halves, each of them consisting of two not quite equal sized parts) is clearly reflected in the dynamic shaping. The amateur (dynamgram in figure 11) is not able to do this: There are generally not as many strong differences in loudness and most of them show only a small extent (the shading lies mainly in the upper area of the dynamram).

Dynagram of a Professional Piano Performance

![Dynagram of a Professional Piano Performance](image)

Dynagram of a Non-Professional Piano Performance

![Dynagram of a Non-Professional Piano Performance](image)

Fig. 10 and 11. Dynagrams of a professional and a non-professional piano performance of Erik Satie's "Gymnopedie No.1" (dark shading: crescendo; middle gray: constant loudness; light shading: decrescendo). The dynamic shaping of the professional is much more strong and extensive. Moreover, his performance reflects clearly the formal structure of the composition.
Discussion
The examples show the possibilities arising from the method introduced here: The dynamic shaping in a piece of music with its whole multidimensionality can be grasped and visualized with the help of a dynagram. Particularly, we get the means for recognizing whether and to which degree a wide shaping of the dynamics exist. A new analytical instrument is put at the disposal of performance research, applicable to all historical and topical performances of which a recording exist. Moreover, the first applications give rise to the supposition that some characteristics of quality in a good performance can be grasped with this new method. We formulate provisionally a hypothesis: The good performance has a wide dynamic shaping, the corresponding dynagram shows intense shading in the lower area. Such a hypothesis goes together with the results of two other studies (Langner & Kopiez 1995, 1996) which deal with the low frequency oscillations triggered by good performances. From this we get the possible application for music education, especially as a result of the option to produce dynagrams on a screen in real time simultaneously to the musician's performance.

Further research is needed to get the bottom of this new method's possibilities. Further hypotheses about the connections with characteristics of a good performance will arise from the application to a broad spectrum of examples, and they will have to be examined. An open question up to this time is, to which of the numerous details visible in a dynagram correlating to human perception exist. Presumably, many of the details will be able to be grouped together and this would lead to the possibility of optimizing the graphic presentation.

A very attractive prospect seems to be the option to apply the procedure not only to loudness curves, but also to the tempo curves well-known in performance research. The algorithms are immediately transferable. Consequently we can get a multidimensional analysis of the musician's tempo shaping and, therefore, have two analogue methods for the two important shaping parameters used by interpreters: dynamics and timing.

Acknowledgement
I would like to thank Reinhard Kopiez (Würzburg) for the many stimulating discussions about this topic and the data of the two Satie-recordings, as well as Rainer Defer (Braunschweig) for the setting-up of the graphs.

References


A RULE FOR AUTOMATIC MUSICAL PUNCTUATION OF MELODIES

Anders Friberg, Lars Frydén and Johan Sundberg

Royal Institute of Technology
Speech, Music and Hearing
S-100 44 Stockholm, Sweden
email: andersf@speech.kth.se, pjohan@speech.kth.se

INTRODUCTION

This is a part of an on-going long-term project which aims at developing a rule system for musical performance. The ultimate goal is to obtain a better understanding of the mechanisms underlying musicians’ transformation of written scores into a musically convincing performance (e.g., Friberg, 1995a; Sundberg, Friberg & Frydén 1991). An important aspect of music performance is the marking of structural units. In a previous article we presented a rule describing how phrases can be signalled in a performance (Friberg 1995b). In this paper we will present a rule which automatically locates possible positions of a lower level boundary, which musicians commonly mark by commas in the score. We will use the term punctuation for the insertion of such commas. They divide the phrases into shorter units, which we, adopting Repp’s terminology (Repp 1992), will call melodic gestures.

The task of the punctuation rule is not only to locate positions of commas but also to signal these commas by micropauses and tone lengthenings in the performance. Here, however, we will focus on the problem of finding appropriate positions for commas.

Many investigations of musical performance are based on an analysis of the musical structure. In this study, however, we aim at formulating the principles that musicians apply when they decide where commas should be introduced in a tone sequence. In other words, our ambition was to elaborate a punctuation rule which is able to synthesize a musically acceptable performance rather than carrying out an analysis of the structure.

There are often several possibilities to place commas. For example, some performers may prefer to direct the listeners’ attention on longer phrases consisting of several gestures while others may prefer to mark each gesture. Also, a given tone sequence may be regarded as beginning with an upbeat by some experts while others regard this upbeat note as a non-upbeat tone, terminating the previous gesture. As a result, the punctuation of different performers may vary considerably. The present rule should therefore be seen as one expert musician’s point of view, leading to a for most listeners acceptable punctuation.

METHOD

The punctuation rule has been developed by means of the analysis-by-synthesis method using the program Director Musices (Friberg, 1995a). The input is the coded music score. The rule only uses information on pitch and duration as specified by the score while other marks such as articulation signs are disregarded. Thus, here we consider the notes the compulsory part of the score information and other signs as performance advices. Such signs have often been inserted by an editor rather than by the composer. To identify appoggiaturas, a phenomenon of great relevance to punctuation, it is necessary to include the harmonic analysis. The program converts such note files into a performance by means of a synthesizer. As in our previous investigations, co-author LF has served as the musical expert providing both musical
judgement of synthesized performances and most of the musical ideas used for elaborating the
rule.

The rule has been developed and tested using a music material of 60 melodic excerpts,
most of which were taken from the classical repertoire, from the Baroque to contemporary
music, but also including some examples of folk music and popular music.

THE PUNCTUATION RULE

The overall strategy of the rule is first to identify potential, weighted locations of commas on
the basis of the melodic context. This is realized by means of subrules which mark all notes
that precede a potential comma. Thereafter, a selection is made among the proposed comma
locations according to the weights and certain other factors.

1. Finding punctuation positions

The subrules, listed below in the order in which they are applied, describe the contexts
associated with location of a potential comma. Figure 1 shows the five main contexts. The
maximum context length is five notes, two notes before and two notes after the current note.
The subrules also add or remove weight values to these locations according to various context
constraints. In cases where a note has received weights from two or more subrules, the
weights are summed. The weight values are used to estimate the importance of the potential
comma. The description of the subrules is complete except for weight values and
mathematical functions.

\[
\begin{align*}
\text{Subrule 1. Before rest} & \quad \text{A note preceding a rest is marked, i. e., the note may be followed by a comma.} \\
\text{Subrule 2. After appoggiatura} & \quad \text{A note terminating an appoggiatura is marked unless it is preceded and followed by a stepwise}
\end{align*}
\]

motion (Fig. 1-1).

\[
\begin{align*}
\text{Subrule 3. Leap} & \quad \text{A note initiating a leap is marked under certain conditions which concern the direction of the}
\end{align*}
\]

pitch changes preceding and following the leap. These conditions also determine the value of

\[
\begin{align*}
\text{Fig 1. Five main contexts generating allocations of potential commas (indicated) according to}
\end{align*}
\]

the punctuation rule. The contexts are (1) after appoggiatura; (2) in melodic leaps; (3) after
the longest note in a series of five; (4) after a note followed by two or more shorter notes of
equal duration; and (5) before a note surrounded by longer notes.
the weight. The weight thus obtained is scaled according to the size of the leap and the
duration of a five note context surrounding the leap (Fig. 1-2).

Subrule 4. Longest of five
A weighted mark is attributed to the center note in a five note context, provided that this note
is the longest in the context. The weight is scaled according to the duration of the long note as
compared to that of its two neighbours (Fig. 1-3).

Subrule 5. Remove marking before rest
A marking occurring on the penultimate note before a rest is removed. The same applies to the
penultimate note of the piece.

Subrule 6. Remove markings of short notes
Marks are removed on all notes with duration shorter than a certain percentage of the average
duration of all notes in the entire melody.

Subrule 7. First short
A note preceding a series of at least two shorter notes is marked. The weight is scaled in
proportion to the duration ratio between the marked note and the first short note (Fig. 1-4).

Subrule 8. Short between long
The first note in a sequence of long-short-long receives a mark (Fig. 1-5).

Subrule 9. Reduce long
This subrule concerns the interaction of the subrules that are triggered by duration contexts,
viz. when a note has been marked by more than by one of the subrules Longest of five, First
short and Short between long. In these cases the smallest contribution to the weight is
subtracted.

Subrule 10. Remove in appoggiatura
A weight mark appearing on the first note of an appoggiatura is removed.

Subrule 11. Remove one in repetition
If two notes equal in pitch both have received weight marks, the smallest of the two weight
marks is removed. If they have equal weights, the first is removed.

Subrule 12. Remove smaller weights
A weight mark is removed if it is smaller than a certain percentage of the average weight
computed over the entire melody.

Summarizing, potential locations of commas appear in two types of contexts,
(1) Contrast in (a) pitch (Leap) and (b) duration. i.e. at note followed by shorter note(s)
(Longest of five, First short and Short between long).
(2) Melodical tension-release (Appoggiatura).
Likewise, two major principles are used for the scaling of weights.
(1) Contrast: The weight is (a) proportional to the duration ratio between the marked note
and a certain context (Longest of five, First short), (b) proportional to the leap size
(Leap), and (c) removed from comparatively short notes (Remove short).
(2) Tempo: Weights are reduced for notes of shorter durations, so that the number of marks
is reduced in sequences of comparatively short notes (Leap).
2. Translating punctuation positions to performance parameters

The remaining positive weight marks are accepted as true punctuations. Such a punctuation is realized in terms of a micropause after the marked note and by a small lengthening of that note's interonset duration. The duration of the micropause and the lengthening are proportional to the note's interonset duration and on a scaling factor $K$, a free parameter for controlling the overall rule quantity. At this stage the weight values are disregarded. A graph of the length and position of the resulting micropauses is shown in Fig. 2.

\[
\begin{array}{c}
\begin{array}{c}
\text{DRO [ms]} \\
\hline
\text{80} \\
\text{60} \\
\text{40} \\
\text{20} \\
\text{0} \\
\end{array}
\end{array}
\]

Fig. 2. Micropauses produced by the Punctuation rule in an excerpt from the string quartet in G major by W. A. Mozart, K387. The upper graph shows the length of the micropause in ms (DRO) i.e., the time interval between the offset of the tone and the onset of next tone.

PRELIMINARY EVALUATION

There are often several possible punctuations, as mentioned above. Therefore, two different measures were used for a preliminary evaluation: (1) the number of cases in which the rule produced commas coinciding with those marked by a professional performer (co-author LF); (2) the number of cases where the rule punctuations were clearly wrong according to the same performer's judgement.

A total of 19 melodic excerpts were selected which exhibited rather complex and in many cases ambiguous phrase structures and hence should serve as a material for a rather "hard" test. The scores of these melodies were printed on paper without any articulation marks. The first task of the subject was to mark by means of commas his preferred punctuations on these scores. In the second task, the scores were marked according to the punctuation rule, and the subject identified the commas that he perceived as musically impossible.

Based on the outcome of the first task, an attempt was made to optimize the weight parameters. By means of an iterative process, each weight parameter was systematically varied so as to minimize the discrepancy between the rule generated punctuations and those of the performer.

Results

Overall, the rule produced 43% more marks than the performer; the total number of the performer's preferred punctuations was 146 while that of the rule generated punctuations was 209. The rule was able to find 76% of the performer's preferred punctuations. The rather large number of rule generated marks that were missing in the performer's versions often reflected the fact that the rule signalled a grouping that differed from that of the performer. In three cases the rule selected an upbeat alternative that was not supported in the musician's version. A total of 86% of the rule punctuations were considered possible according to the second test.
The optimization process improved these numbers to some extent. The rule then produced only 25% more marks than the performer but still identified 70% of his preferred punctuations. Of the rule generated punctuations 90% were considered acceptable by the performer. Thus the optimization eliminated 30% of the impossible rule generated punctuations.

Our tentative conclusion is that the rule mostly succeeded in producing an acceptable distribution of commas and generated a clear majority of the punctuations marked by our single musician subject.

DISCUSSION

The grouping of notes and its realization is by no means obvious to an expert musician. This process involves most features in the music such as tempo, note values, pitches, harmonic progression as well as emotional and other intentions, of which some are not apparent in the musical surface (Berry, 1989). The results so far surprisingly show that a rather simple input, essentially consisting of pitches, durations and rather short contexts, are indeed sufficient for predicting a majority of the commas marked by a professional musician. To improve the agreement between observations and predictions in this respect a more complex input is likely to be required, e. g., expanding the context, using the harmonic analysis more extensively. Also it would be worthwhile to introduce a procedure for identification of sequences, although inspection of LF’s markings revealed that the punctuation was not necessarily exactly copied throughout the entire series of similar gestures. A minor modification of the gesture may call for a different punctuation. In some cases the rule managed to produce such variations of the punctuation, since an exact copy of the punctuation is granted only when pitch and duration sequences reappear verbatim.

This is a preliminary investigation using a rather small number of excerpts and comparatively simple rules. Given these limitations, the results were surprisingly promising. Thus, it seems worthwhile to expand the scope by matching a larger material of excerpts against a greater number of professional performers’ preferences. Also, it would be instructive to attempt an optimization of the weight parameters, e. g. by means of further use of an iteration technique and also by using neural networks (Bresin, in preparation).

ACKNOWLEDGEMENTS

This work was supported by the Swedish National Council for Research in the Humanities and Social Sciences and by the Bank of Sweden Tercentenary Foundation.

REFERENCES

On the technology of interpretation

Andranik Tangian
FernUniversität, Feithstr. 140, D-58084 Hagen, Germany
Andranik.Tangian@FernUni-Hagen.De

Abstract

Musical performance is considered as falling into two levels, interpretation and nuancing, the former requiring mainly analytical, the latter technical skills. Interpretation is understood as finding a structure in a musical text by its segmentation and determining the relation between the segments. Nuancing is understood as a means of showing the segments by intonation patterns.

We develop a system of rules to perform segments and provide an example of using this analysis-based method while interpreting the First Movement of Vivaldi’s “Spring” from *Four Seasons*. We describe its interpretation realized with a computer and the one by a violin pupil appropriately instructed.

KEYWORDS: Analytical interpretation, segmentation, phrasing, intonation patterns, expressive pulses, music teaching.

1 Introduction

Following our earlier studies on music interpretation (Tanguiane 1992a–b) and their extension to theater (Tanguiane 1996, 1997), we develop an approach to interpretation as finding a structure in a score by segmentation and showing this structure in playing. The background segmentation is partially rule- and partially heuristics-based (Lerdahl and Jackendoff 1983; Schaffer 1995).

It is supposed that a segment is shown by an ‘intonation wave’, or ‘intonation envelope’ in dynamics and timing (Todd 1985; Clarke 1988). To be perceived, it should have clear borders and distinguishable phases of intensification, culmination, and relaxation.

The musical structure has several levels of segmentation. At the lowest level, the melodic line is divided into motives which are smallest meaningful segments, each having a unique stress (Thomassen 1982; Drake & Palmer 1993). The intonation envelope is determined by dynamics and tempo deviations from given nominal values, according to a certain pattern (Parncutt 1994). The deviations from the nominal values for every motive are regulated by patterns of the superior segmentation level, corresponding to phrases, each having a unique intonation peak said to be the target. Such a hierarchical grouping of segments with unique ‘centers of gravity’ has to be realized up to the level of musical form.
This intonation pattern hierarchy is similar to Clynes’ (1977, 1983, 1995) expressive pulses, understood as a hierarchy of patterns of regular deviations from nominal time and dynamics values. Following studies by Repp (1989, 1990) and Thompson (1989), our approach differs from that of Clynes in at least two points: We suppose that (a) the pulses depend on the context rather than on the style of a given composer, and (b) they are carriers of structural rather than emotional information.

To provide experimental evidence to our arguments, we consider the First Movement of Vivaldi’s “Spring” from *Four Seasons*. Its interpretation has been realized

(a) by an eight-year old violin pupil taught to distinguish phrases by dynamics, timing, and articulation;

(b) with a computer program with dynamics and timing specifications derived from the segment structure of the piece.

2 Model

We consider music as creating some special time arrangement by means of acoustical events. This means that events constitute time patterns, transformations of which determine transformations of musical time. For instance, repetitive rhythmic patterns are used as reference units for tempo tracking; melodic waves characterize phrase structure, etc.

The acoustical events are multiattributed (number of pitches, timbre combinations, durations, etc.). Each attribute can develop independently of others, implying a rather complex arrangement of musical time.

In our study, the task of interpretation is understood as realizing these temporal processes in a logically consistent way. An emotional performance has been attempted next at this advanced starting point.

We shall refer to three levels of segments which are distinguished with regard to their length, and the type of their ‘centers of gravity’ shown by intonation.

The intonation is a means to show temporal connectedness and tension between acoustical events, and, on the other hand, to oppose and to rank the groups of events. The intonation is formed by pitch, dynamics, timing, and articulation.

Note the ‘objective’ nature of the intonation determined by the pitch contour given in the score. Using execution nuances is aimed at either enhancing the pitch contour, or at a certain contrast to it. This can be done in several alternative ways which is the task of interpretation.

A motive is a shortest meaningful melodic segment with a single intonation stress. A motive is analogous to an indivisible speech expression which may contain several words but only one stress. In poetry/vocal music a motive corresponds to a line, or a part of a line which can be singled out.

A phrase is a union of several motives with a single intonation target. In poetry/vocal music a phrase usually corresponds to a text line whose ending is the intonation target.

A period is a union of several phrases with a single intonational culmination and a single intonation resolution. In poetry/vocal music a period corresponds to a rhyme with
a rhythmically and intonationally equilibrated pattern.

Note that a motive is characterized by a single-peaked intonation envelope. A phrase has a multiple-peak envelope dominated by the target. A period is characterized by an envelope with several waves but a unique culmination.

Generally speaking, a hierarchical organization of information is rather universal. For instance, a broadcasting wave is a carrier of the acoustical wave which is its envelope. In turn the acoustical wave is a carrier of spectral information, the spectrum is a carrier of formants, etc.

Similarly, an intonation wave is a carrier of information of a higher structural level which is determined by relationships between intonational peaks, peaks of groups of peaks, etc, thus determining the time envelope of musical form. Since we consider musical performance as an art of time arrangement, the intonation waves are regarded as a carrier of time marks whose multidimensional structure constitute the meaning of music.

The accent, target, culmination, and resolution are all hierarchically organized 'gravitation points' of musical drive. To show that drive, dynamics, timing, and articulation (timbre) are used.

3 Computer simulation

The computer simulation has enabled to specify the nuances and principles of their applications. It seems that our methodology can be applied to playback animation option to music score editors similar to experiments by Honing (1990) and Mazzola (1994).

In computer modeling we have applied micro and macro-level rules.

Micro-level deal with motives and short phrases where performance is based mainly on metrical cues. Metrical cues have been used to sustain metrically strong notes (up to 110% of nominal values (more suspension sounds as exaggeration), and enhancing them with dynamical stress +3dB.

At the macro-level (phrases longer than 1 measure) we have applied the following rules:

1. Long ascending sequences of transposed patterns are performed with gradually increasing tempo up to 105% of the initial value, with increased dynamics, with the goal being the last metrically accentuated note.

2. Descending sequences of transpositions are performed diminuendo. with a tempo decrease.

3. A literal repetition is performed with abrupt piano as echo and with no tempo change.

4. Motive and phrase boundaries are marked by a tempo decrease. The higher the segment stands in the segment hierarchy, the more salient the tempo variation. Consequently, the ending has the greatest final retard (in our experiments 60% of initial tempo).

5. Each entry of the theme is marked by the tempo reset to its initial value.
The experiments with computer show that long intonation waves need less salient envelopes than short intonation waves. It looks as if the perception becomes more refined over larger data windows. We can draw analogy to the Fourier transform where lower time resolution enables to obtain finer frequency resolution.

As a general remark, after having applied the rules mentioned, the computer playback has been no longer mechanical and, after adjustments, has become quasi-emotional. The segmentation has been done manually, but it can be done semi-automatically by known segmentation rules (Lerdahl and Jackendoff 1983). In our experiments we have used only rhythmical cues specified in detail in (Tanguiane 1993, Chapter 6).

We are far from thinking that our or similar rule-based system can run automatically. In most cases manual adjustments are highly desirable. However, we are almost sure that the number of basic execution rules is not numerous and most of them can be implemented in a ready-to-use routines.

An interesting problem is to specify nuances with a few parameters which could be adjusted by a potential user or selected from a menu. For example, in our experiments the tempo curve has been evaluated from the segment length (longer segments require less tempo increase if the tempo increase should not be perceived). In this case the parameter is: Timing animation (not perceptible, with tempo change paramenter calculated automatically), or Tempo change (with a quantitative adjustable parameter).

4 Teaching experience

The approach described has been tested in teaching an eight year old violin pupil. The pupil has been explained that making an interpretation is like building from a Lego construction set with elementary bricks of different shape and color. The ‘shape’ corresponds to phrases of certain resemblance, and ‘color’ is the execution technique.

The idea of intonation envelopes has been explained both verbally and illustrated by playing. In particular, all three types of performance of the three types of segments mentioned has been distinguished. Next, the rules described have been applied to different pieces of music in order to illustrate their general meaning. One of difficulties in children’s teaching turned out to perform long nuances, as homogeneous crescendo with tempo increases. Another difficulty turned out to be the feeling of initial tempo (for tempo reset at the theme entries). The ensemble playing with the teacher providing the piano accompaniment has been used to fix the main ideas of interpretation.

The exercises have been aimed at both getting the desired dynamics and timing over phrases, and at acquiring a feeling of the musical form in general. Besides improving technical skills, and learning the text, a pupil has been faced to the problem of achieving clearly specified sound. The exercises have become more creative, more purpose-oriented, and more motivated. As a side effect, the efforts to get the nuances required have had a positive effect on the sound quality in general.

As a result, a meaningful interpretation of the piece, rather large and complex for a young musician, has been achieved.
5 Conclusions

Let us recapitulate the main statements of the paper.

1. Interpretation is understood as segmentation and determining the accents within the segments. Musical form is shown by a hierarchy of intonation waves in dynamics and timing. This hierarchy determines the performance structure as a system of successive targets.

2. Segments of different types (e.g., relating to the theme, or to a transient passage) have distinctly different intonation patterns. Dynamics and timing are used consistently to show tension and relaxation (e.g., crescendo with accelerando).

3. Computer experiments enable to judge about sufficiency/insufficiency of the set of performance rules used. These rules have been also used in teaching as a base of free performance, not restricting emotional freedom. Rational explanations make the interpretation task well defined, and the training becomes more purpose-oriented.

4. The idea of Clynes’ expressive pulses is revised as related to the segment structure rather than to the metric structure of a piece.

References


Author index

Abrams, R.M. 684
Agmon, E. 122
Ahlbück, S. 263
Aksnes, H. 551
Allyn, D.A. 706
Altenmüller, E. 669, 675
Anagnostopoulou, C. 655
Appletos, L.J. 471
Ash, R.L. 257, 613
Ashley, R. 499
Atalay, B.V. 206
Baroni, M. 484
Behne, K-E. 675
Belardinelli, M. Olivetti 71, 82, 87
Bellugi, U. 681
Berggren, U. 517
Bergman, D. 430
Bertrand, D. 397, 407
Blumenberg, H. 286
Bochkarev, L.L. 153
Boeckkarev, L.L. 625
Brodksky, W. 170, 317, 323
Burnsed, V. 435
Calabrese, I. 87
Cambouropoulos, E. 533
Camiz, P. 82, 97
Caterina, R. 484
Chen–Haiteck, L. 412, 418
Chin, C.S. 105
Chouvel, J-M. 545
Chronopoulou, M. 222
Ciardi, F. Cifariello 91
Clarke, E.F. 471, 589, 594, 702
Clough, J. 128
Cohen, D. 605
Cordes, I. 235
Costa–Giomi, E. 446
Cross, I. 407
Davidson, J. 329
Deliègue, I. 48, 407
Descombres, V. 111
Di Matteo, R. 76
Dibben, N. 594
Djef, M. 577
Drake, C. 286, 465
Dry, A. 475
Edlund, B. 140, 697
Erola, T. 377
Eichert, R.G. 527
Eitan, Z. 275
Engelbreitsen, N. 128
Ericson, G. 571
Eriksen, A. 165
Finnès, L. 635
Folkestad, G. 147
Franek, M. 298
Friberg, A. 582, 719
Frydén, L. 719
Fuller, R. 706
Fung, C.V. 619
Furneaux, S. 210
Fyk, I. 440
Gabrielsson, A. 430, 475, 479, 650
Gál, B. 504
Gerhardt, K.J. 684
Ginsborg, J. 189
Goday, R.I. 557
Good, J.M.M. 329
Griffiths, S.K. 684
Gruln, W. 669
Gudmundsdottir, H.R. 424
Hargreaves, D. 147
Henik, A. 170
Hettergott, A. 371
Hodges, D.A. 33
Hoshino, E. 333
Huang, X. 684
Imberty, M. 13, 23, 61
Jens, J.J. 688
Juslin, P. 493
Järvinen, T. 157
Jørgensen, H. 171
Kleinen, G. 644
Klempe, H. 563
Kruger, M. 706
Krumhansl, C. 3
Krystecki, J. 440
Lado–Bordowsky, Y.A. 357
Lagerlöf, I. 577
Lammers, M. 706
Lamont, A. 403, 407
Land, M.F. 210
Langford, G. 684
Langner, J. 713
Lehiste, I. 241
Lehmann, A.C. 165, 629
Leleu, J-L. 206, 365
Lerdahl, F. 45
Levittin, D.J. 117, 681
Lidman–Magnarsson, B. 339
Lindström, B. 147
Lindström, E. 292
Littleon, D. 457
Madison, G. 302
Mamalis, N. 392
Mancia, E. 82
Martinez, I.C. 269
Mates, J. 298
McNeil, A. 195
Mondry, H. 605
Moreno, M. 111
Nattiez, J.-J. 55