Music Perception
Winter 2004, Vol. 22, No. 2, 297–350

Influences of Large-Scale Form on Continuous Ratings in Response to a Contemporary Piece in a Live Concert Setting

STEPHEN McADAMS
Institut de Recherche et Coordination Acoustique/Musique (STMS-IRCAM-CNRS) and Ecole Normale Supérieure

BRADLEY W. VINES
McGill University

SANDRINE VIEILLARD & BENNETT K. SMITH
STMS-IRCAM-CNRS

ROGER REYNOLDS
University of California at San Diego

Listeners responded continuously at the world and North American premiere concerts of *The Angel of Death* by Roger Reynolds using one of two rating scales: familiarity or resemblance of musical materials within the piece and emotional force. Two versions of the piece were tested in each concert in different presentation orders. Functional data analysis revealed the influence of large-scale musical form and context on recognition processes and emotional reactions during ongoing listening. The instantaneous resemblance to materials already heard up to that point in the piece demonstrates strong relations to the sectional structure of the music and suggests different memory dynamics for different kinds of musical structures. Emotional force ratings revealed the impact of computer-processed sounds and a diminution in emotional force with repetition of materials. Response profiles and global preferences across the two versions are discussed in terms of the multifaceted temporal shape of musical experience.

CREATORS in the temporal arts—music, cinema, dance—often devote a great deal of energy and imagination to the formal aspects of their works. Given that such forms can extend over several minutes, several tens

Address correspondence to Stephen McAdams, CIRMMT, Faculty of Music, McGill University, 555 rue Sherbrooke ouest, Montréal, Québec, Canada H3A 1E3 (e-mail: smc@music.mcgill.ca)
ISSN: 0730-7829, electronic ISSN: 1533-8312.

of minutes, and in rare cases several hours, one wonders at times what the subjective reality of such forms might be if one takes into account human memory limitations. The form in which one poses the question of form is already rife with presumptions about what it is or might be, and the role it plays in the temporal arts.

The present project sought to address some of these issues in the case of a long piece of contemporary music composed by the last author especially for the project, in interaction with the psychologist authors. Experiments were carried out under ideal conditions for ecological validity at the premiere concerts of the piece in Europe and North America using a continuous response method. This may well be the first time such an experiment has been performed in such natural conditions for concert music. The piece was conceived to have formal properties that allowed for the study of certain dynamic processes in listening, all the while having to fulfill entirely the artistic goals of the composer. In collaboration with the composer, it was decided to focus on two aspects of listening experience: one related to perceptual processing of musical structure over time, in particular the sense of familiarity or resemblance of materials that are repetitions or variations of previously heard materials; the other related to the emotional force felt by the listener as a function of the musical structure. Online response protocols were used to measure listeners' perceptions and reactions in the concert hall. The project raises questions concerning the nature of musical form, the role of familiarity and recognition in music listening, evolving musical emotions, and the analysis of all of them with continuous response data.

Musical Form and Temporality

Notions of musical form can be roughly classed into three fuzzy categories with unclear boundaries. One class considers form "out of time" and might be considered an architectonic approach. This approach is classic in many Western music theoretic traditions. It involves descriptions in terms of sections (sonata form, ritornello form, rondo form, theme and variations form, etc.) or of nested hierarchies as in the work of Schenker (1935/1979) and Lerdahl and Jackendoff (1983) with hierarchies that go all the way up to the entire piece, even if that piece lasts an hour and a half, as in the case of a Mahler symphony. Memory and the temporal aspects of listening processes are not often taken into consideration in these approaches. Indeed one might say that they implicitly presume memory to be infinite and exhaustive. In their defense, it should be acknowledged that with a few exceptions, their aim is not to describe the *experience* of the form but some more or less objective (or "neutral") level of the musical *object* under study.

A second approach to form is basically to deny its psychological existence. One radical instantiation of this approach is the concatenationist theory of Levinson (1997), which considers the experience of music to exist in the moment and the overall experience of a work to be the concatenation of successive moments: "Musical form is centrally a matter of cogency of succession, moment to moment, part to part" (p. 14). This concept has resonance with Stockhausen's notion of *Momentform* as embodied for example in his piece *Kontakte* (1959–1960) for electronic tape and percussion: "Each moment, whether a state or a process, is individual and self-regulated, and able to sustain an independent existence. The musical events do not take a fixed course between a determined beginning and an inevitable ending, and the moments are not merely consequents of what precedes them and antecedents of what follows; rather the concentration on the Now—on every Now—as if it were a vertical slice dominating over any horizontal conception of time and reaching into timelessness, which I call eternity: an eternity which does not begin at the end of time, but is attainable at every moment" (Wörner, 1963/1973, pp. 46–47). The striking feature of this conception of form is the absence of memory or perhaps a denial of its implication in accumulating musical information, inferring temporal trajectories, and comparing things heard now with things heard in the not-too-distant past in the piece. At the same time, Levinson (1997) seems to accept that the past can influence the present through "tacit, unconscious correlation of present passages or bits with earlier ones, rather than explicit, conscious grasp of relationships of a broad-span sort" (p. ix). So while adopting a concatenationist stance in contradistinction to the explicit consideration of large-scale architectonics, he nonetheless allows for memory-like processes to operate implicitly. One might conclude that the distinction he really wants to make is between implicit and explicit apprehension of form.

A third approach is to consider form to exist "in time." In this conception, large-scale form is the shape of experience through time and its resonating reminiscences, rather than a structure out of time that one holds before the mind's ear in its entirety. Such is the conception of the experience of form of composers like Roger Reynolds (1987), who considers that the work of the composer is to shape experience through sound. The temporal component of the experience is thus crucial. This approach evokes the newly emerging domain of cognitive dynamics (Ward, 2002) of which Jones (1976; Jones & Yee, 1993) is one of the pioneers by bringing considerations of time and temporally bound processing into the realm of musical and everyday listening. One of the keys of this approach is the interaction among processes through time such as perceptual organization, attention, memory storage and retrieval, and emotional, affective, and aesthetic reactions. Because such processes are dynamic, their study suggests the use of methods that are sensitive to temporal evolution in the face of changing information in order to address perceptual, cognitive, and emotion-related issues in music listening and the resulting psychological experience. It also suggests the importance of using full, real pieces in naturalistic settings.

Many of the important questions about the dynamics of musical listening experience remain unposed. What are the relative contributions in different kinds of music of processes of perceptual processing, of attentional dynamics, or the interplay between memory and perceptual processing in the sense of an unfolding apprehension or comprehension of music, in the evolving affective and aesthetic reaction to the music? What is the nature of what remains when the music is over? What are the time constants of various aspects of the memory of the musical experience? Do some decay more rapidly than others? What are the relative contributions of the musical architecture and the performance characteristics of its realization to the temporal form of the experience? The present study examined the contribution of familiarity and recognition processes and evolving emotion to music's temporality.

Familiarity and Recognition

One question that was addressed in the present study was the perception of musical similarity or the recognition of associations between original musical materials and variations or transformations of those materials during listening to the piece. How does the sense of evolving familiarity or recognition shape one's "comprehension" of a musical work? The sense of familiarity, of having already heard something related, is something that composers play on intuitively a great deal with thematic presentation and development, and also at more global levels with the structuring of musical sections on the basis of differences in texture, harmony, register, and instrumentation. Such processes contribute not only to the sense of changing from one section to another in a piece, but also to the associative links across the piece that are established by recognition.

In the study of perception of musical structure and structural relations, most work has focused on comparisons outside of musical contexts: melody recognition or discrimination paradigms (DeWitt & Crowder, 1986; Dowling & Bartlett, 1981) or judgments of musical similarity in paired comparisons (Matzkin, 2001; McAdams & Matzkin, 2003) are good examples. Few studies of complete musical works have been performed. Of those that have, many have used a segmentation task (at times with post-performance descriptions of the form) to explore the processing of larger scale musical structure (Aiello, 1994; Berz & Kelly, 1998; Clarke & Krumhansl, 1990; Deliège, 1989; Karno & Konečni, 1992; Pollard-Gott, 1983). Globally, the results show effects of change in texture, tempo, density, register, instrumentation, or pauses at section boundaries. But Ayari and McAdams (2003) showed a finer grained understanding of modal organization in improvised Arabic music, at least in Arab listeners, who possess the necessary cultural schemata.

Dowling, Kwak, and Andrews (1995) and Dowling, Tillmann, and Ayers (2001) used an adaptation of the continuous running-memory paradigm (Shepard & Teghtsoonian, 1961) to study the fluctuating nature of what is remembered about a musical "event" over time. The change over time of the effectiveness of different kinds of retrieval cues for melodic excerpts results in the evolution of melody recognition: decreases in performance are found for some kinds of comparisons, increases for others.

Tulving's (1983) theory of episodic memory and a consideration of a shift in retrieval strategy from short- to long-term memory are cogent here. Traces of past events are stored in memory via a process of encoding. Memory is queried when a retrieval cue is presented. Information from the cue is combined with information from the trace, resulting in what Tulving refers to as "ecphoric" information. Recoding with the ecphoric information can change the memory trace, and this information provides a basis for the memory system's response to the query. Changes in the effectiveness of various cues in bringing about recall over time occur, suggesting a critical role of recoding in the dynamics of memory operation in a continuous activity such as reading or listening to music.

Consideration of these processes of interaction between memory coding and perception leads one to conceive of the evolving sense of familiarity or resemblance one experiences over time in a large-scale musical work, when several themes are presented in repetition or in various kinds of transformations or variations. Their temporal organization creates a fluctuating sense of the association of the present with what has been heard in the near past. Along these lines, Krumhansl (1998) used continuous measures of memorability to explore the role that musical topics play in the experience of music by Mozart and Beethoven and showed the topics in Mozart to be crucial to the definition of musical form.

Musical Emotion

The study of emotional responses to music is undergoing great expansion at present. The interests range from characterizing the emotions felt in listening to music, to pinpointing what aspects of musical structure generate them, to the major dimensions of emotional experience such as arousal or activity and valence, to the way they evolve over time (cf. Bigand, Vieillard, Madurell, Marozeau, & Daquet, submitted; Dalla Bella, Peretz, Rousseau, & Gosselin, 2001; Juslin & Sloboda, 2001; Schubert, 2001). It is their evolution over time that is of primary interest to the present study.

Krumhansl (1998) used continuous ratings of the amount of emotion, openness, and memorability in pieces by Mozart and Beethoven with the aim of studying the relation of musical structure to the concept of intonation units in Chafe's (1994) theory of the temporality of discourse. She found that new ideas were introduced at points of low tension and neutral tempi, and musical tension tended to increase to a peak just before a segment ending. Sloboda and Lehmann (2001) used a continuous response technique to explore how the intensity of a simple emotion is affected by interpretative decisions during the course of performances of relatively short pieces of music. They found effects of musical structure and performance variations on the emotionality profiles: peaks and troughs of emotional experience were associated with structural features of music, as had been found previously (Krumhansl, 1996; Palmer, 1996; Panksepp, 1995; Sloboda, 1991).

Vines, Nuzzo, and Levitin (submitted) used continuous ratings of musical tension as a measure of musical affect. They concentrated on derivatives of the temporal response profiles to characterize what they termed affective velocity and affective acceleration in relation to expectation and release in music in order to capture the dynamics of musical tension. They interpreted their data as indicating that affective energy is stored and released as musical tension increases or decreases, creating an arch structure of musical tension and release, which recalls the breathing of music and changing emotional dynamics proposed as a crucial component of musical meaning by Meyer (1956).

Time Series and Functional Data Analysis

Online approaches to probing the psychological dynamics of perception, cognition, and emotion of music are being used increasingly in the field to measure perception and recognition processes and emotional reactions. A number of continuous tracking devices have been used in the study of musical emotions and musical tension: a spring-loaded squeezing device (Nielsen, 1983), a dial (Frederickson, 1995; Madsen & Fredrickson, 1993), and a slider on a computer screen (Krumhansl, 1996, 1997; Krumhansl & Schenck, 1997; Vines, Wanderley, Krumhansl, Nuzzo, & Levitin, 2003). Other online techniques involve pressing keys or foot pedals to indicate segmentations in an ongoing musical piece (Ayari & McAdams, 2003; Clarke & Krumhansl, 1990; Deliège, 1989). In one case, correlations were performed between musical tension and a number of physiological measures (Krumhansl, 1997).

Using continuous response protocols raises a number of methodological issues concerning the continuous rating method itself, the characterization and analysis of response profiles, and the use of inferential statistics in their interpretation. One question that arises with continuous response measures is reliability. Schubert (1999) measured test-retest reliability using his two-dimensional emotional space interface after delays of 6–12 months and found a moderate correlation ($r = .74$). He also emphasized the importance of stimulus coding at some level in establishing relations between response profiles and the music to which a listener is responding (Schubert, 2001). On the basis of correlational analyses and measurements of response delays, he suggested that, at least for emotional response, a sampling rate on the order of 2 Hz is sufficient. Concerning the possibility that such continuous measurement techniques might interfere with the musical experience, Madsen and Coggiola (2001) found that post-performance global aesthetic ratings of a 20-min excerpt from Puccini's *La Bohème* were higher when listeners manipulated a dial to make instantaneous aesthetic ratings during listening than when listeners just made the global judgment after passively listening. This result suggests a positive effect of the attention that is required to perform such a task.

The emerging field of functional data analysis (FDA), pioneered by Ramsay and Silverman (1997, 2002), has addressed the issue of continuous data analysis and interpretation. In this approach, statistical analysis techniques normally applied to scalar data (single values) are extrapolated to functional data such as temporal response profiles. The assumptions and interest of this technique are described in detail in the Ramsay and Silverman books and in more summary fashion in Levitin, Nuzzo, and Ramsay (submitted) with applications to music psychology in Vines et al. (submitted). An important feature of FDA is to be able to specify precisely in time when important subjective reactions are taking place with respect to the musical stimulus or when two sets of functions related to the same musical structure, but measured under different conditions (e.g., different performances) differ significantly from one another. Given the interest of studying effects of large-scale form on instantaneous responses, implicit and online, and potential differences between interpretations of the piece at its two premieres, we will use aspects of these FDA techniques for both rating scales.

The Piece *The Angel of Death*

There are several important structural aspects of *The Angel of Death,* for piano solo, chamber orchestra, and computer-processed sound (Reynolds, 2001), to be examined in relation to real-time responses. It is of great import to music psychology to work with real pieces of music that have a compositional structure that is known in detail and that allows for testing certain hypotheses concerning musical form. The structure of the piece (Fig. 1) will be only briefly recalled here because it is described in detail in Reynolds (2004) and is analyzed by musicologists Philippe Lalitte and François Madurell in a series of chapters in McAdams and Battier (2005). A number of abbreviations will be used for the various sectional elements in the piece, which are summarized in Table 1. *The Angel of Death* is based on five thematic materials (T1 to T5), which vary in duration from about 30 s to 2.5 min in the performances under consideration. Each theme is composed of several subsections, one of which is considered to be central to the temporal trajectory of the theme: the core element (gray rectangles in Fig. 1; see Lalitte et al., 2004). In addition to the themes, there are other developmental sections in which transitions between thematic materials occur (TR13, TR24) or in which thematic materials are combined in parallel (CB24, CB35, CB123). There are also a large nondirectional, nonthematic region called Other, a section of repetitive ostinatic strata (REP), and a Piano Interlude (PnoInt).

The piece is organized in two parts that traverse these materials in different ways, but have similar global temporal structures in that the theme core elements, Other, REP, and PnoInt occur at identical positions in the two. This temporal structure is shown by the superposition of the two parts in Figure 1. Around these temporal pillars, the materials are developed differently. In the Sectional (S) part, there are clear section boundaries between themes, Transition and Combination regions, and the other types of regions. Two of the Combination regions (CB35 and CB123) do not occur in this part. In the Domain (D) part, the materials are more interpenetrating and overlapping, giving a more organic and diffuse sense of the musical flow. Note for example that the difference between the clear delineation of T1, T2, TR13, and T3 in the S part is lost with the overlapping materials from T1–T3 in the same temporal region in the D part. This distinction between the clear sectional structure of S and the more organic flow of D is one of the crucial aesthetic aims of the composer. The materials that are played by the piano in one part (e.g., T1, TR13, REP in S and T2, CB24, TR24, T5 in D) are instantiated by the orchestra in the other part. So although the thematic core elements return at the same moment in both parts, they change from one instrumentation to the other. Two regions stay in the piano in both parts: Other and PnoInt.

The piece can be played in two versions depending on the order in which the S and D parts are played: S-D or D-S (see Reynolds, 2004). The computer-processed sound, organized into 10 sections that start at specific points in the instrumental score, always occurs in the second part of the piece. It starts just before the end of the first part, occupies a solo bridge

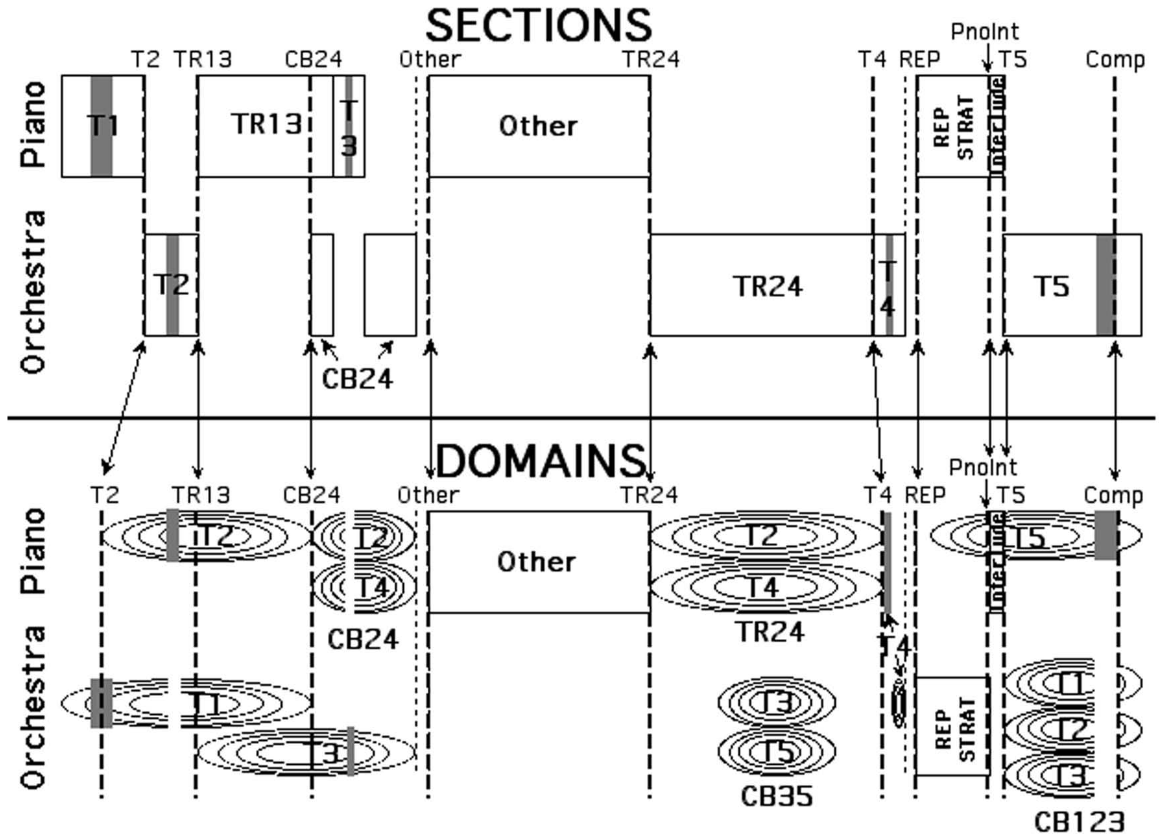

Fig. 1. Structure of the two parts (Sections and Domains) of *The Angel of Death* and structural boundaries (dashed vertical lines) used in statistical tests. (See Table 1 for the key to abbreviations.) Note that the Combination regions CB24, CB35, and CB123 in the D part are indicated by the superposition of the corresponding thematic materials. The latter two do not have analogues in the S part. Gray boxes indicate core elements of the themes. The labels for the structural boundaries just above the dotted lines are used in the text.

TABLE 1
Abbreviations for Sections of The Angel of Death

Abbreviation	Description	Type
T1 – T5	Themes 1–5	Section
TR13, TR24	Transition regions progressing from materials of one theme to those of another, e.g., from Theme 1 to Theme 3 for TR13	Section
CB24, CB35, CB123	Combination regions containing materials from several themes, e.g., Themes 2 and 4 for CB24	Section
REP	Region of repetitive, ostinatic strata	Section
PnoInt	Piano interlude	Section
D1, D2, D5, D6, D9, D10	Computer "images" of the domain type	Section
S4, S7, S8, S11	Computer "images" of the sectional type	Section
S	Sectional part of the piece	Part
D	Domain part of the piece	Part
S-D	Version with S part followed by D part	Version
D-S	Version with D part followed by S part	Version
DSA1	Version D-S performed 1st in Paris concert	Performance
SDA2	Version S-D performed 2nd in Paris concert	Performance
SDU1	Version S-D performed 1st in La Jolla concert	Performance
DSU2	Version D-S performed 2nd in La Jolla concert	Performance

region between the two parts lasting a little over 2 min (D10), accompanies the second part (D6, D5, D2, S11, S4, S8, D9), replaces Other with a computer-transformed version in the second part (D1), and occupies a solo coda region after the end of the second part that lasts nearly 2 min (S7). The piece ends with a brief, reflective piano Epilog.

Aims of the Study

The structure of the piece raises a number of interesting issues concerning online processing of musical materials and musical form that form the basis of our exploratory hypotheses and questions. The composer conceived the S and D parts as having very different characters, and hypothesized that the more abrupt changes in material in the S part would give a more clearly sectional character than would the overlapping, interpenetrating deployment of thematically similar materials in the D part. Further, given that the piece has two versions, differing in the order of the parts, questions concerning the influence of previous hearing of one part on the perception of the other can be addressed: that is, does the large-scale context affect the perception and recognition of musical materials and listeners' emotional reactions to them? Also, because the computer layer always accompanies the second part, comparing the same part in two versions allows us to study the influence of this layer on the instrumental material (albeit confounded with the part's position in the piece). Finally, in the two premiere concerts, both versions were played, but in different orders so that longer-term influences of previous hearing of the music could be studied (although these are confounded with differences in interpretation and cultural milieu, as well).

To address these issues, two continuous rating scales were used: one that concerned the perception of musical materials and structure (familiarity or resemblance of what is currently being heard compared to anything heard from the beginning of the piece) and another that addressed listeners' emotional responses (emotional force felt in response to the music). The two measures draw on markedly different perceptual, cognitive, and affective operations. The familiarity/resemblance scale is purely dependent upon memory at different levels of processing. For example, a segment may be familiar because surface features (texture, timbre, pitch range . . . any feature that can be extracted from the musical sound or visual experience) are shared with another segment in memory. In addition, higher level processing might be involved for the memory of relations between parts over longer time spans. This latter kind of processing would be involved in the recognition of certain kinds of transitions or contours of change—anything involving change over time, in fact. Because there are different levels of processing, different time scales might be involved in the sense of familiarity with the materials or their resemblance to materials encountered earlier in the piece.

The emotional force rating involves very different capacities for the listener. It references the listeners' emotional responses, not their memory traces, although memory might potentially influence emotion. Emotional responses are related to physiological mechanisms (Hatfield, Cacioppo, & Rapson, 1994; Krumhansl, 1997) and they may have valence—a tendency to induce attraction or repulsion. However, following from Schubert (2001) and from the instructions given to listeners, it is likely that the unidimensional emotional force scale employed taps into the arousal or activity level component of emotion rather than the valence component. One criterion for selecting a unidimensional emotional force scale is that we were more interested in the evolution of the force of the emotional reaction over time and its relation to the musical structure than with the actual categories of emotional experience that might be evoked and which may vary widely from one listener to the next. Another criterion was simply the financial and technical factors related to the fact that we had to manufacture 128 response devices and administer them in real time in a live concert setting.

Method

The Angel of Death was performed in its two versions at the premiere concerts in Paris and La Jolla, CA. The world premiere concert took place within the Agora Festival of the Institut de Recherche et Coordination Acoustique/Musique (IRCAM) on June 7, 2001 in the 448-seat Grande Salle of the Centre Pompidou. Approximately 350 people attended the concert, which was performed by the Ensemble Court Circuit, conducted by Pierre-

André Valade, with Jean-Marie Cottet playing piano. The North American premiere took place within the Time Forms Festival on April 17, 2002 in the 785-seat Mandeville Auditorium at the University of California at San Diego in La Jolla, CA. Approximately 550 people attended the concert, which was performed by the SONOR Ensemble, conducted by Harvey Sollberger, with Gloria Cheng playing piano.

PARTICIPANTS

Paris Concert

One hundred six listeners participated in the experiment. They were given free entrance to the concert in exchange for their participation. Forty-nine performed familiarity ratings and 57 performed emotional force ratings. Of the 49 listeners performing the familiarity rating task, 33 returned the questionnaires at the end of the concert: 17 males and 16 females, aged from 20 to 67 years ($M = 38$, $SD = 14$). Thirteen classed themselves as nonmusicians, 16 as amateur musicians with from 4 to 22 years experience playing music ($M = 11$, $SD = 6$), and 5 as professional musicians with from 8 to 40 years experience ($M = 18$, $SD = 13$). Thirty-five listeners performed the task for both versions of the piece, 5 performed it only for the S-D version, and 9 for the D-S version only. Of the 57 listeners performing the emotional force rating task, 39 returned the questionnaires: 22 males and 17 females, aged from 20 to 69 years ($M = 36$, $SD = 14$). Eighteen classed themselves as nonmusicians, 17 as amateur musicians with from 4 to 22 years experience playing music ($M = 10$, $SD = 2$), and 4 as professional musicians with from 12 to 15 years experience ($M = 14$, $SD = 1$). Forty-one listeners performed the task for both versions of the piece, and 16 performed it only for the D-S version.

La Jolla Concert

Ninety-nine listeners participated in the experiment. They were given free entrance to the concert in exchange for their participation. Forty-six performed resemblance ratings, and 53 performed emotional force ratings. Of the 46 listeners performing the resemblance rating task, 44 returned the questionnaires: 31 males and 13 females, aged from 19 to 66 years ($M = 35$, $SD = 15$). Ten classed themselves as nonmusicians, 21 as amateur musicians with from 4 to 20 years experience playing music ($M = 9$, $SD = 5$), and 13 as professional musicians with from 8 to 43 years' experience ($M = 27$, $SD = 15$). Twenty-seven listeners performed the task for both versions of the piece, and 17 for the S-D version only. Of the 53 listeners performing the emotional force rating task, 51 returned the questionnaires: 31 males and 20 females, aged from 13 to 81 years ($M = 37$, $SD = 19$). Nineteen classed themselves as nonmusicians, 23 as amateur musicians with from 2 to 52 years of experience playing music ($M = 11$, $SD = 4$), and 9 as professional musicians with from 8 to 25 years of experience ($M = 16$, $SD = 6$). Thirty-three listeners performed the task for both versions of the piece, 3 performed it for the D-S version only, and 16 performed it only for the D-S version.

STIMULI

The stimuli consisted of live performances of *The Angel of Death*.[1] For each concert, both versions (S-D and D-S) were played. In the Paris concert, the D-S version (duration 34 min 24 s) was played first, followed by the first two movements of an electroacoustic piece by the same composer, *Versions/Stages* (1986–1991), an intermission, the last two

1. The two versions of the piece (S-D and D-S) recorded at the Paris premiere will be released on an audio CD to be distributed with the e-book on the Angel Project (McAdams & Battier, 2005). Study scores of *The Angel of Death* are available from C. F. Peters Corp. which publishes Reynolds's music.

movements of *Versions/Stages,* and then the S-D version (duration 34 min 14 s). In the La Jolla concert, the S-D version (duration 33 min 45 s) was played first, followed by an intermission, a silent film by Jeff Perkins, *Sam Francis: To Paint is to Forget,* accompanied by two chamber pieces by Morton Feldman, *Durations I* (1960) and *Durations IV* (1961), a second intermission, and then the D-S version (duration 34 min 22 s). Real-time continuous ratings were requested of listeners only for the two versions of *The Angel of Death.*

PROCEDURE

Listeners were ushered into the hall and handed an instruction sheet explaining the use of the real-time response device, a biographical questionnaire, an informed consent form, and an additional questionnaire to be filled out after the concert concerning their impressions of the experiment and the piece.

A brief presentation of the Angel Project and the place of the in-concert experiment in it was given, followed by an explanation of the aims of the experiment and instructions. Particular emphasis was placed on the fact that the listeners were being asked to listen to and experience the piece, to monitor a specific aspect of their responses to it, and to translate that response into a position of the slider on the box, continuously through time. In the La Jolla concert, the explanation session took place an hour before the concert so that nonparticipants would not have to sit through this explanation. As such, the participants were able to practice the use of the response boxes while listening to the first 6 minutes of a recording of another piece by the composer (*Summer Island* for oboe and computer-processed sound, 1988). Due to various administrative and cultural constraints, this was not possible at the Paris concert, so nonparticipants were present during the briefer verbal presentation of the experiment.

Each participant held a response box equipped with a slider connected to a potentiometer. The response boxes came in two colors: beige for the Emotional Force scale and black for the Familiarity or Resemblance scale. For each numbered box, the name of the scale appeared at the top and the end points of the sliders were labeled (Fig. 2). Slightly different versions of the scale related to recognition of musical materials were used in the Paris and La Jolla concerts. In Paris, the scale was called *Nouveauté/Familiarité* (Novelty/Familiarity, upper left panel, Fig. 2). Remarks by participants and subsequent laboratory experimentation led us to think that the notion of familiarity might be confused with a general sense of familiarity informed by the listener's past experience, rather than that specifically generated by materials in the piece being heard, despite explicit instructions to relate it only to the piece being heard. For

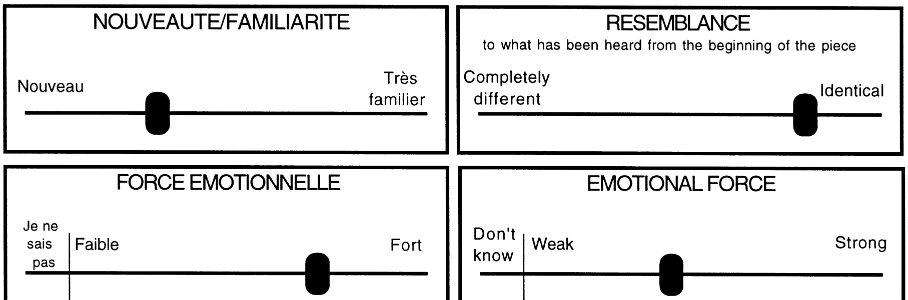

Fig. 2. Response box labels for the Paris (right) and La Jolla (left) concerts for the Familiarity or Resemblance scales (upper) and the Emotional Force scale (lower). The vertical line across the slider track in the Emotional Force scale indicates a slight resistance to passing into the "Don't know" zone.

La Jolla, therefore, the scale name was changed to “Resemblance to what has been heard from the beginning of the piece” (upper right panel, Fig. 2). The response boxes were mechanically modified for the Emotional Force scale. On the far left was a zone labeled “Don’t know” (lower panels, Fig. 2), which was distinguished tactilely by a slight resistance in moving the slider into this zone (the slider having to move past a rubber washer inside the box). This tactile cue was to avoid having listeners inadvertently place the slider in this zone. The decision to include the “Don’t know” zone was based on pilot experiments in which it seemed that having this possibility reduced experimental noise. The periods in which the slider was in this zone were treated as missing data in subsequent analyses.

Instructions for the Familiarity and Resemblance scales explained that listeners were to judge how familiar what they were hearing at the present instant was to anything they had heard from the beginning of the piece (Paris), or how much it resembled anything heard from the beginning of the piece (La Jolla). Emphasis was placed on using the full scale and on rating varying degrees of familiarity or resemblance in a continuous fashion.

Instructions for the Emotional Force scale (Paris and La Jolla) invited listeners to rate continuously the force of their emotional reactions, regardless of the nature and positive or negative valence of those reactions. In other words, it was the strength or magnitude of the emotion that was being sought. They were informed that if they really didn’t know what they were feeling, they could place the slider in the far left zone, but that they should minimize the use of this zone. Again emphasis was placed on continually monitoring their reactions and moving the slider accordingly.

Listeners participating in the experiment were placed in the center of each auditorium in order to benefit maximally from the six-channel spatialization of the electroacoustic part of the piece. Seats with Familiarity/Resemblance and Emotional Force response boxes were interspersed so that neighboring listeners were not rating the same scale. In all, there were 64 response boxes for each scale, although not all of them were used.

The following labels will be used to distinguish the versions of the piece, the concert venues, and the order in which a given version was heard in the concert (see Table 1). “DS” refers to the version with the Domain part followed by the Sectional part, and “SD” refers to the reverse configuration. “A” refers to the Agora Festival concert in Paris and “U” to the UCSD concert in La Jolla. The numbers 1 and 2 refer to the order in which the piece was heard in a given concert. So DSA1 concerns the D-S version played first in the Agora Festival concert.

The individual temporal profiles were examined and the data for some listeners were rejected on the basis of criteria that sought to select listeners who remained active throughout the whole piece and used the scale according to instructions. Profiles were rejected if at least one of the following characteristics were observed: (1) if the rating started above (i.e., to the right of) the middle of the scale (familiarity/resemblance only), (2) if there were long periods of several minutes in which the profile remained flat, (3) if the profile showed only a steady increase (familiarity/resemblance only), (4) if the profile showed only a binary alternation between the top and bottom of the scale with no graded variation, (5) if the profile remained within a very small range throughout the whole piece (emotional force only), (6) if the profile evolved extremely slowly (slow variations over 5–10 minutes; emotional force only), and (7) if the response behavior consisted of rapid jitters with direction changes every 2–3 seconds that are unlikely to reflect the time course of emotional response to music (emotional force only). Table 2 lists the number of profiles rejected in each category for each performance. While it may be that some listeners with long flat periods did indeed experience the piece in this way, we decided to remove them in the doubt that they had actually stopped responding. Our analyses focused on how the response profiles of people who responded actively to the piece corresponded to the structure of the piece. We felt that this focus justified the selection criteria used. All unrejected data are included in the analysis without reference to musical training due to the small sample sizes for some groups.

TABLE 2
Number of Response Profiles Rejected and Reasons for Rejection for the Different Performances

	Familiarity		Resemblance		Emotional Force				
Reason for rejection	DSA1 $N = 40$	SDA2 $N = 44$	SDU1 $N = 46$	DSU2 $N = 27$	DSA1 $N = 57$	SDA2 $N = 41$	SDU1 $N = 50$	DSU2 $N = 36$	Total $N = 341$
Begins right of middle	8	8	0	0	—	—	—	—	16
Flatliner	5	6	4	2	11	8	2	3	41
Steady rising	2	2	0	0	—	—	—	—	4
Binary behavior	0	2	2	2	0	0	0	1	7
Minimum range	—	—	—	—	0	3	3	0	6
Extremely slow variation	—	—	—	—	0	7	0	2	9
Jittery	—	—	—	—	0	0	1	0	1
Percent rejection	37	41	13	15	19	44	12	17	25

NOTE—*N* refers to the total number of participants.

EQUIPMENT

The equipment used in the Paris concert was shipped to La Jolla for that concert. The only difference was that the French labels were replaced by English labels (Fig. 2). The range of the slider in the response box corresponded to a continuous voltage between 0 and +5 V. Sets of 16 boxes were connected with XLR connectors to a concentrator box from which a cable carried all 16 analog signals. Two concentrator boxes were connected to an AtoMIC Pro converter (http://www.ircam.fr/produits/technologies/atomic/specs.html) with DB25 connectors. The AtoMIC Pro converted the analogue voltages of the 32 input channels into MIDI messages, linearly mapping the 0–5 V range onto the 0–127 discretized MIDI scale. Four such AtoMIC Pros were used to convert the signals from the 128 response boxes. The outputs of the AtoMIC Pros were sent in parallel to two Emagic Unitor 8 MIDI concentrators, which interleaved the MIDI messages and sent them to a Macintosh G4 computer. In the Macintosh, the incoming MIDI signals were recorded using Digital Performer (v. 2.7), synchronously with a mono audio feed from the mixing table in the concert hall so that the MIDI data could be temporally aligned with the audio signal from the performance. All MIDI and audio recordings were made in parallel on two Macintosh computers as a safety precaution. Full details of the equipment setup are available on http://www.ircam.fr/pcm/bks/death/.

When the voltage from a potentiometer changed (indicating a change in the position of the slider), the AtoMIC Pro constructed a MIDI control-change message, where the controller number identified which potentiometer had moved and the value corresponding to the voltage (and thus the position of the slider). The frequency of messages from each individual slider was limited to a maximum of 10/s. The data were stored as time-tagged values. These series were treated as step functions and were subsequently sampled at 2 Hz for the data analyses. The sampled data will be referred to as response profiles or temporal profiles.

Familiarity/Resemblance Scale

RESULTS

The raw temporal profiles were linearly scaled so that individual profiles filled the interval between 0 and 1. With the familiarity and resemblance data, we were interested in both the slider displacement profiles and the first derivatives of the profiles, that is, the rate at which the displacement changed over time. To compute the derivatives and have them in an interpretable form, we needed to fit smooth functions to the raw profiles. We used functional data analysis (FDA) techniques to do this (Levitin et al., submitted; Ramsay & Silverman, 1997), using Ramsay's (2003) FDA software package written in MatLab (The Mathworks Inc., 2003). Listeners' individual profiles were converted into functional objects by approximation with a basis expansion, which was composed of a set of concatenated third-order polynomials (B-splines) that were continuous in the displacement function and in the first two derivatives at the junction points, called knots. For our data, 600 B-spline segments were used for the full piece, and 300 were used for the S and D parts when analyzed separately. This means that the data were regressed onto polynomial segments over time spans of approximately 3–4 s, which gives a good fit to the data. The resulting functional objects are considered to approximate the smooth continuous process underlying the recorded data samples. The rms error between the raw data and the functional data profile was computed. The mean rms errors for our four datasets (DSA1, SDA2, SDU1, DSU2) vary from .019 to .028, with a global rms error representing about 2.3% of the full scale. All displacement and derivative data reported below relate to the functional data profiles derived from the familiarity or resemblance rating profiles.

Different performances of the same version of the piece do not have identical timings. Therefore, when profiles are compared across performances they need to be aligned temporally, a process called registration. We used a landmark-registration technique in which various points (landmarks) in the piece were aligned, and the intervening regions were linearly dilated or compressed in time accordingly. For full-piece comparisons, 38 points were used corresponding to beginnings of the major sections described above and the core elements. For comparisons involving S or D parts, 18 of these points were used.

To perform inferential statistics on comparisons between performances, a one-way between-subjects analysis of variance (ANOVA) was performed at each time sample (a functional F test). The analysis outputs the F statistic at each sample, and the F values are displayed as a function of time. This approach allows one to visualize the evolution of the statistical difference between two populations. Such a procedure raises of course the issue of multiple comparisons, an area that has not been addressed much in the functional data literature. As such we decided to adopt a primarily qualitative approach to interpreting the statistical profiles at this early stage in the development of these techniques.

The mean scaled profiles for the two versions of the piece in the two concerts are displayed in the upper panels of Figures 3 (S-D version) and

4 (D-S version). These data are very rich, and only a few select points will be considered here. There is a global trend toward increasing familiarity/resemblance ratings as the piece progresses, with a fair amount of fluctuation occurring as the musical materials start to be encountered in varied forms. Note, for example, that the initial version of Other (a long, nondirectional, contemplative region with a fairly homogeneous texture) tends to increase over its time course, whereas the second occurrence (in its computer-processed version, D1) is higher. Another example is the increase in ratings for the piano interlude from the first to the second part and extending to the Epilog, for which the Interlude serves as a kind of premonition. In the mean profiles, there are sudden changes at many points indicating a relative consensus among listeners that familiarity or resemblance changed in a given direction. In many cases, these sudden changes occur on or just after sectional boundaries in the piece (shown by the dotted vertical lines in Figs. 3 and 4), indicating a collective sensitivity to the macrostructure of the piece. There are some boundaries that are stronger than others in the sense that more abrupt changes in the mean profile are found at them. We will return to this point in more detail below in an analysis based on the derivatives of the profiles.

Another feature that is apparent from these figures is that the version that was heard second (thicker line) tends to have a higher mean profile than does the version heard first (thinner line). There appear to be many larger scale structural similarities between the two concerts for a given version (the profiles have globally similar shapes), although there are many differences at the level of local details. Finally, there are considerable differences in shape between the same part (S or D) depending on whether it occurs in first or second position in a given version, e.g., compare the S part from T1 to D10 in the first part of SDA2 (Fig. 3) to the S part from T1 to S7 in the second part of DSA1 (Fig. 4). These various differences will be examined below with reference to the inferential statistics.

Examination of the details of the mean profiles reveals certain aspects that may be related to the use of the scale by the participants. At times a sudden decrease occurs at section boundaries, even though the material has been heard before in the first part or in the other version, but then the profile rises rapidly as the listeners recognize the material. One example of this effect is found at the beginning of Other in the first half SDA2 and at the beginning of D1 (the computer version of Other) in the second half of SDU1 (Fig. 3). So the scale is used to rate local material contrast as well as a more general familiarity or resemblance to material heard previously. In this sense, it involves different kinds of memory processes: detecting local differences between what is being heard and what was still in working memory, but also similarities between materials separated in time by way of long-term memory. The relative frequency of change in the individual profiles suggests that with a few exceptions listeners based their

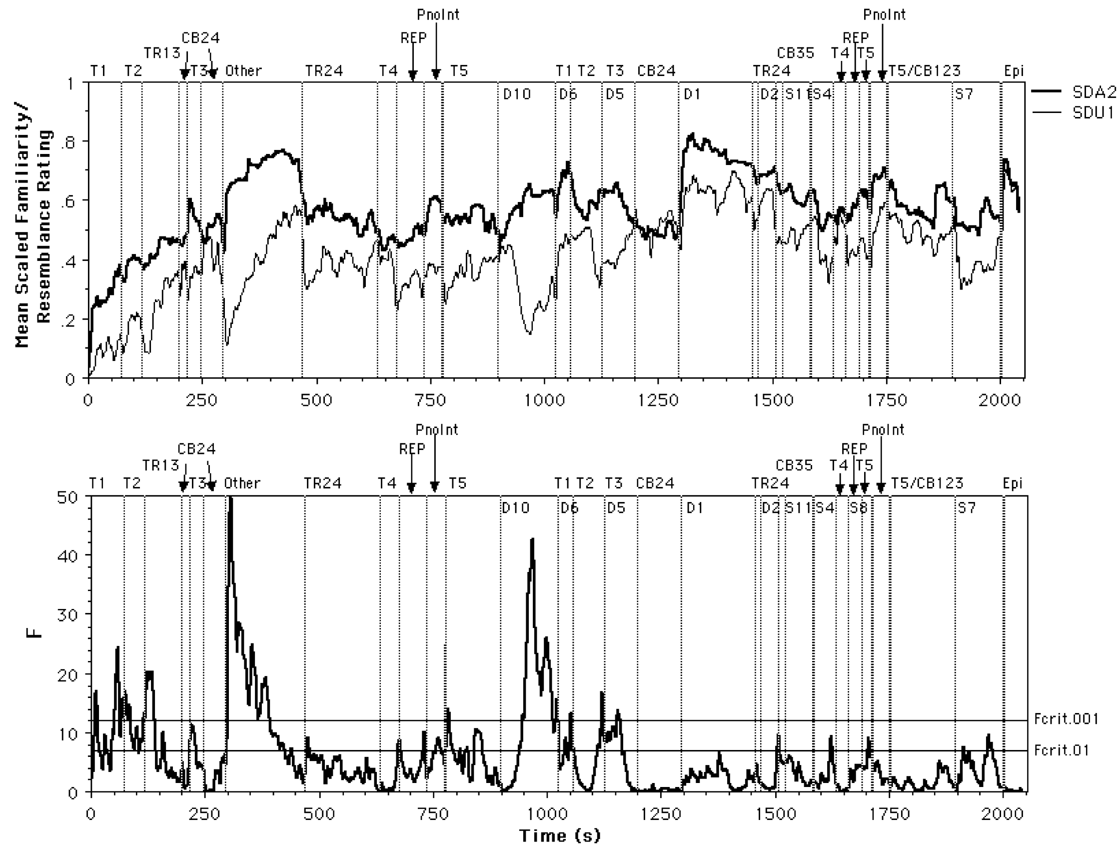

Fig. 3. Mean scaled Familiarity and Resemblance profiles for the S-D version of *The Angel of Death* and the functional *F* test representing their statistical comparison. See Figure 1 for the section labels. The sections labeled D1, D2, S4, D5, D6, S7, S8, D9, D10, and S11 correspond to the computer "images." Vertical dotted lines in the upper panel represent sectional boundaries in the piece (see Fig. 1). The S part goes from T1 through T5 in the first half and the D part from T1 through T5/CB123 in the second half. Horizontal lines labeled Fcrit in the lower panel represent significance criteria at p = .01 and .001.

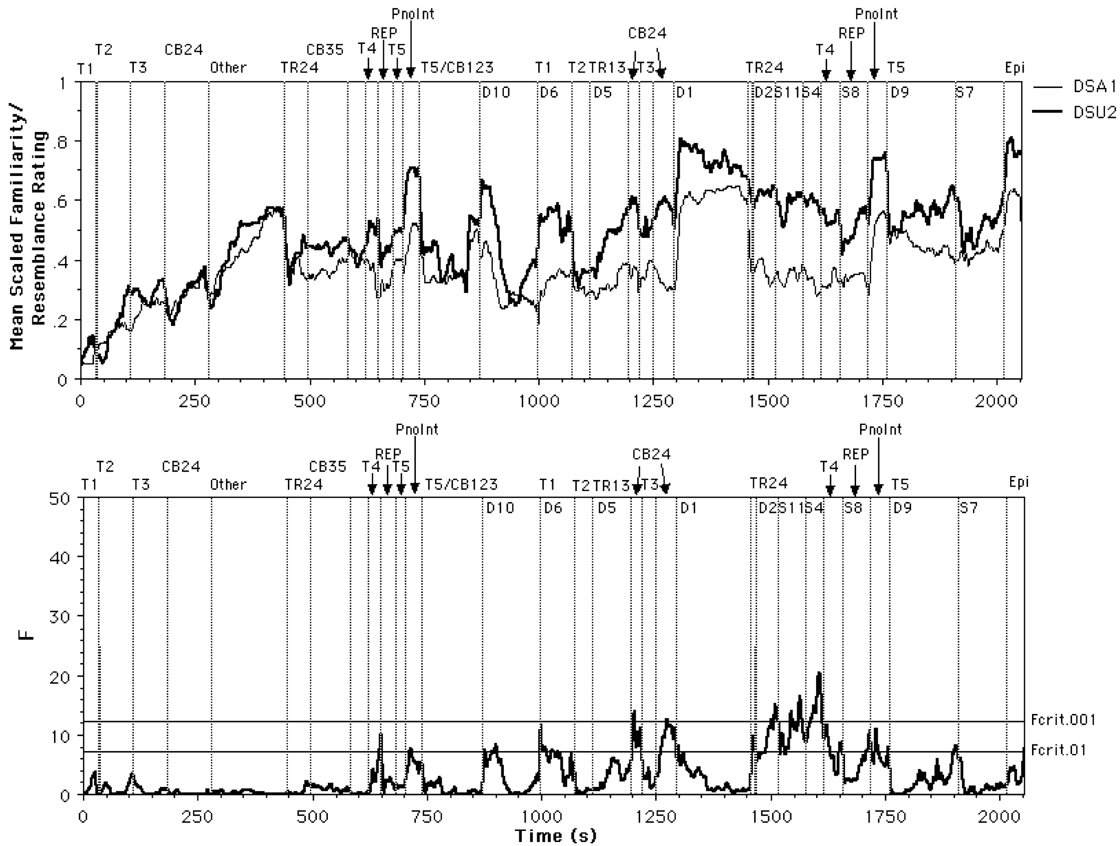

Fig. 4. Mean scaled Familiarity and Resemblance profiles for the D-S version of *The Angel of Death* and the functional *F* test representing their statistical comparison. The D part goes from T1 through T5/CB123 in the first half and the S part from T1 through T5 in the second half. (See Fig. 3 caption.)

ratings on material differences at the sectional level rather than at the level of specific melodic or rhythmic patterns. The relative complexity and density of the piece and the extended nature of the thematic materials themselves may have biased listeners' response behavior in this direction.

Comparisons between Concerts

We can compare the entire mean (temporally aligned) rating profiles across concert venues to see a combined effect of the order in which the version was heard and the different performance styles and venue (concert hall, cultural milieu, language, etc.).[2] The functional *F* test gives us an indication of where the two sets of data diverge significantly from one another, and we can base inferences on that. In Figures 3 and 4, the lower panels show the functional *F* test, the structure of the piece, and the value of *F* at the two criterion levels. There is a high degree of resemblance between the Paris and UCSD data sets for D-S versions (Fig. 4), but the profiles are less similar for the S-D versions (Fig. 3). This is revealed by the greater extent of significantly different regions in S-D: 506 s total vs. 273 s total for D-S at $p = .01$ (246.5 s vs. 57.5 s at $p = .001$), and the fact that the correlations between profiles are .82 for D-S and .69 for S-D. There is a general tendency for the version heard second to have a higher mean profile (SDA2 > SDU1, DSU2 > DSA1), but this is particularly strong for the S part, whether it is in first position (S-D) or second position (D-S). So we already see that the effect of the order in which the versions are heard is different for the two parts, i.e., we have a differential effect of large-scale form on instantaneous ratings as a function of the nature of the musical instantiation of the materials (sectional or domain). These observations are supported by the increasing number of significantly different regions revealed by the functional *F* test in the S parts compared to the D parts in both versions.

Further, differences between profiles related to the order of hearing of a given version (e.g., SDU1 vs. SDA2 in Fig. 3) often appear as a greater descent in the first version heard at the arrival of new material, or a greater rise in the second version or second part heard at the recognition of returning material.

---

2. Ideally we would have liked to have had a fully factorial design with factors Venue (and Interpretation), Version, and Order of presentation of the versions. We attempted to do both orders of presentation of the versions (S-D/D-S and D-S/S-D) at the Paris concert on successive days. Unfortunately, only a small number of people attended the first concert and only a small proportion of those performed the experiment. The number of listeners was insufficient for the data analyses in the first concert. We were not able to do two concerts with counterbalanced orders in La Jolla. We therefore do not have complete independence of the three factors. Indeed each factor is the interaction between the other two, e.g. Venue = Version × Order.

Comparisons of the Same Part across the Two Versions

We compared the S and D parts among themselves to look at combined effects of order (heard first or second), position (occurring in first or second part), and electroacoustics (with or without the computer-processed sounds), e.g., the S part in SDA2 (first position without electroacoustics in the version heard second in the concert) vs. the same part in DSA1 (second position with electronics in the version heard first). Globally, the version heard second always has a higher mean rating than the one heard first for the S part: SDA2 > SDU1 and DSU2 > DSA1. However, the picture is different for the D part. The mean ratings are always higher when the D part occurred in the second half of the piece: SDA2 > DSA1 and SDU1 > DSU2. In both cases, the other factor (Position for S and Order for D) tends to have an effect as well, judging by the relative positions of the profiles. This result suggests very different kinds of memory processing for the two parts: D is more affected by shorter-term contextual factors, and S is more affected by longer-term factors that persist through another piece and the intermission within the whole concert context.

Note that if the computer-processed layer had a strong effect on familiarity or resemblance ratings, one would expect a decrease along these scales when it was present: S in D-S would have lower ratings than S in S-D, and D in S-D would have lower ratings than D in D-S. However, the differences always go in the direction of higher ratings on a second hearing, whether the part being examined was heard first or second in the presence of the computer layer. This result suggests that the instrumental part dominates these ratings.

Implicit Detection of Sectional Boundaries

The first derivative of the scaled displacement function was used to examine changes in the state of familiarity or resemblance that would indicate implicit detection of a section boundary. The functional data profiles were first smoothed to remove some of the small ripples at corners of the step functions (known as Gibbs phenomenon) resulting from the spline regression. Functional data analysis smoothing is accomplished by applying a roughness penalty to the data: a smoothing parameter, λ, controls the trade-off between accurately approximating the actual mean of the data and limiting variability in the curve (Ramsay & Silverman, 1997, chap. 4). When $\lambda = 0$, the resulting curve will be identical to the actual mean curve for the data. As λ increases, the resulting curve will approach the standard linear regression fit to the data. For the derivative analysis of the present data set, we used $\lambda = 5$. The first derivatives of the smoothed individual profiles were then computed and scaled by their standard deviations. This derivative function gives a peak when there is a sudden change in the position of the slider. Since we are interested in any sudden change irrespective of its direction, the absolute value of the scaled derivative function was determined. We wanted to determine if the structural form and the boundaries therein caused state changes in the perceived familiarity of the music. Certainly, we lose information about increasing and decreasing levels of familiarity in this analysis, but those characteristics of the data are addressed in other analyses presented later. We established a threshold value equal to two standard deviations from the mean of the derivative function for each participant, and values below this threshold were set to zero to reduce noise in the data due to small fluctuations in slider position. Figures 5 and 6 display the mean thresholded absolute scaled derivative functions (subsequently referred to simply as derivative profiles) for the two concerts, respectively, with vertical dotted lines indicating the major sectional boundaries in the piece. It is quite clear from these figures that the majority of the peaks in the derivative profiles correspond to sectional boundaries, with a few exceptions that include the beginnings of core elements of T2 and T5. Not all sectional boundaries are responded to, however. It should be noted at the outset that these conceptual boundaries at the level of the organization of regions drawing from different thematic materials are not all conceived as breaks in the musical discourse by the composer and are often quite different in their realization in the S and D parts.

In order to test the perceptual strength of section boundaries that have analogues in both S and D parts, we developed a measure of boundary strength on the basis of the derivative profiles. Tested section boundaries are shown as thick dotted lines in Figure 1. The measure of boundary strength was the mean of the 5-s period starting at the boundary minus the mean of the 5-s period ending at the boundary. The stronger the perception of the boundary, the higher this measure should be. For each of the four performances, a repeated-measures ANOVA was performed on Part (2) $\times$ Boundary (10) with boundary strength as dependent variable. The Geisser-Greenhouse adjustment was applied as necessary and epsilon and the corrected probability are cited where appropriate. The main effect of Boundary was significant for all performances, SDA2: $F(9,189) = 3.2$, $p = .012$, $\varepsilon = 0.512$; DSA1: $F(9,252) = 11.9$, $p < .0001$, $\varepsilon = 0.634$; SDU1: $F(9,351) = 7.6$, $p < .0001$, $\varepsilon = 0.679$; DSU2: $F(9,198) = 6.3$, $p < .0001$, $\varepsilon = 0.604$. Boundaries that were strong across parts and performances included Other, TR24, PnoInt, and T5. For the four performances, only the D-S versions had significant Part $\times$ Boundary interactions, DSA1: $F(9,252) = 2.8$, $p = .019$, $\varepsilon = 0.572$; DSU2: $F(9,198) = 2.4$, $p = .032$, $\varepsilon = 0.626$. These interactions are due to a divergence in boundary strength between parts. For DSA1, they diverge for CB24 and Other (with significant positive strength for S and near-zero strength for D) and for Comp (with positive strength for D). For DSU2, the strength for Other is higher for S, whereas that for Comp is higher for D (the entry of the D10 computer section in the D part created more change than did S7's entry in the S part). There are more significant boundaries in the D-S version than in the S-D version, and more in the S part than in the D part.

The effect of a part's position and of the order of hearing a given version were tested across concerts and versions in a mixed ANOVA, with Boundary (10) as a repeated measure and Order (2), Position (2), and Part (2) as between-subjects variables.[3] The Part × Position interaction, $F(1,220) = 15.2$, $p < .0001$, shows that S and D parts have mean strengths across the 10 boundaries that are equivalent when they are in first position. However, the strengths increased in second position for S and decreased for D (Fig. 7a). So S's boundaries are more salient after having heard D, whereas D's boundaries are less salient after having heard S. The triple Part × Position × Boundary interaction, $F(9,1980) = 3.4$, $p = .0013$, $\varepsilon = 0.798$, shows that Position interacts in a complex way with Part in its effect on boundary strength (Fig. 7b,c). Note that there is a tendency for S to have greater boundary strengths in Position 2 (D-S version) and for D to have greater strengths in Position 1 (D-S version as well). This result suggests potential effects of both large-scale form or accumulated experience with the materials and style on the parsing of a musical structure at the sectional level.

Tests on the Implicit Recognition of Core Elements of the Themes

Poulin-Charronnat et al. (2004) showed that musician and nonmusician listeners had difficulty recognizing thematic materials from *The Angel of Death* when the instrumentation changed from piano to orchestra or vice versa. However, the experimental task was performed outside of the musical context within which those materials are normally embedded. One might imagine either that the musical context could reinforce the memory traces and lead the listeners into a strategy of listening that would enhance recognition, or that the intervening materials would make it even more difficult to recognize the thematic materials when they return in a different instrumentation much later (≈17 min). We examined the implic-

3. We do not have actually have a pure between-subjects design, since some (but not all) participants performed the task for the two versions from which are derived Position and Order. A full mixed ANOVA with only the participants that provided data for both versions was performed: between-subjects factor Venue (2), repeated measures on Version (2), Part (2) and Boundary (10). There was only one minor difference in a two-way interaction with this subset of participants compared to a between-subjects analysis with all of the participants. So we decided to proceed with a between-subjects analysis in order to be able to use all of the data.

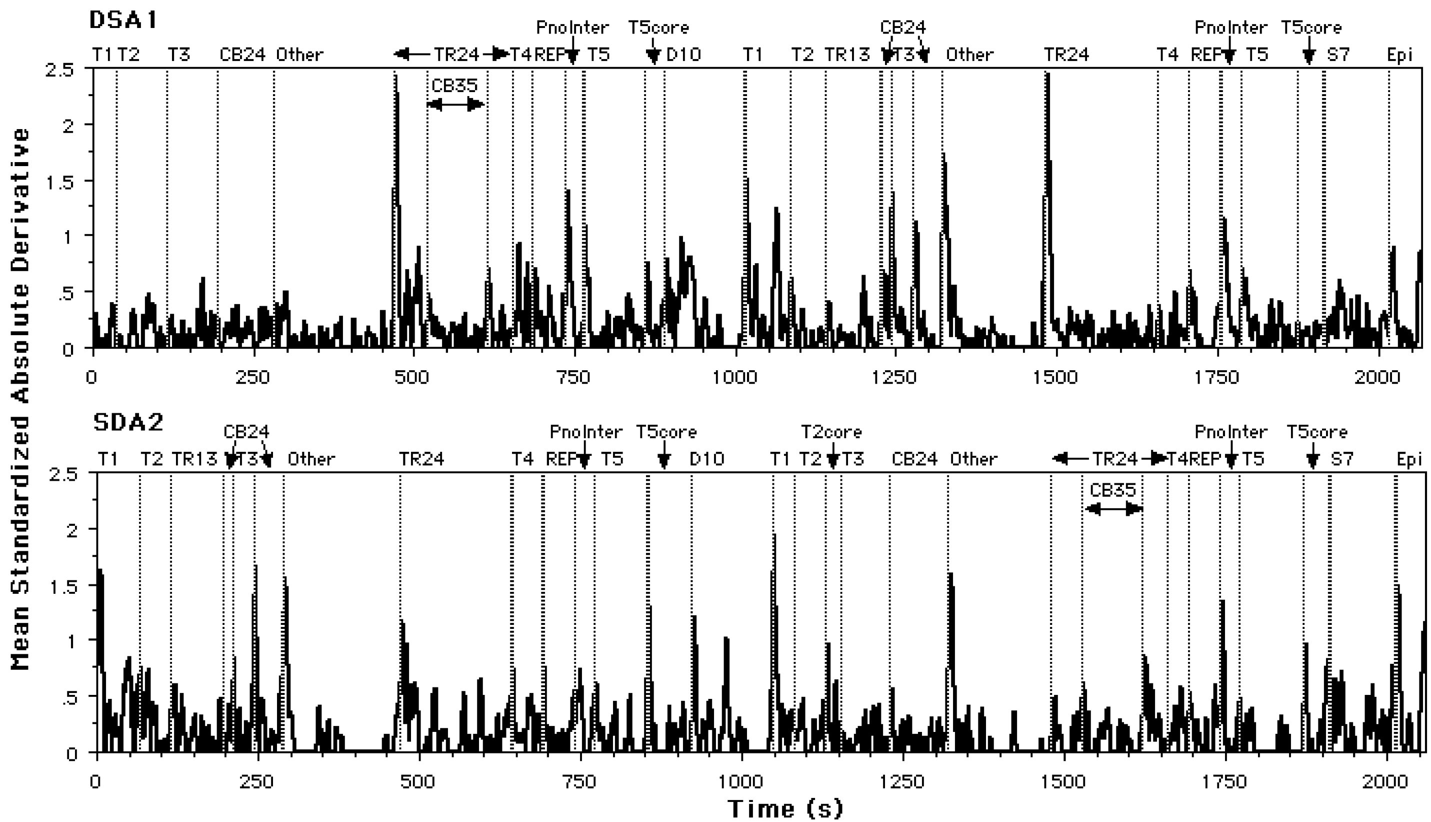

Fig. 5. Mean standardized absolute derivative of the Familiarity profiles for D-S (upper panel) and S-D (lower panel) versions in the Paris concert. Vertical dotted lines in the upper panel represent sectional boundaries in the piece (see Fig. 1).

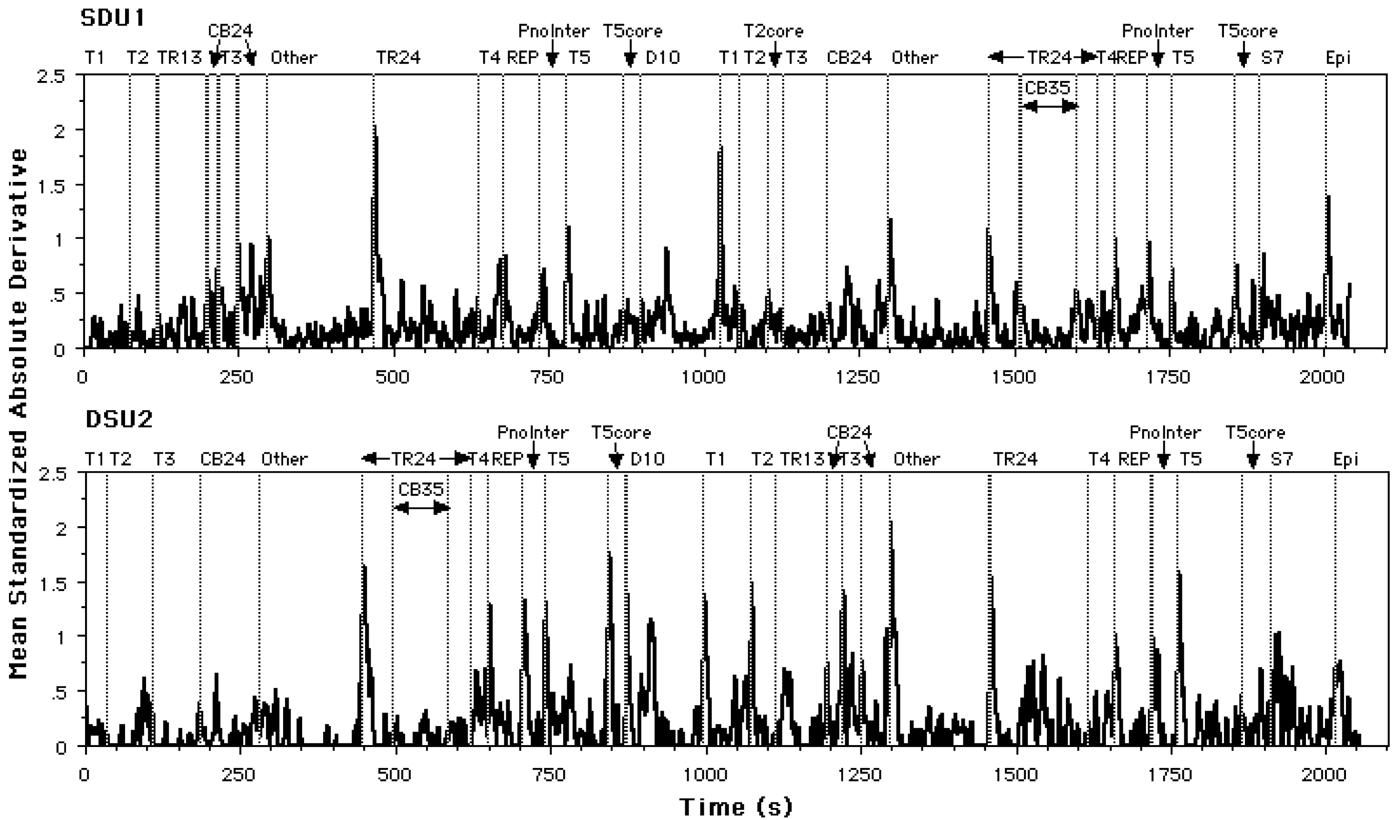

Fig. 6. Mean standardized absolute derivative of the Resemblance profiles for D-S (upper panel) and S-D (lower panel) versions in the La Jolla concert. Vertical dotted lines in the upper panel represent sectional boundaries in the piece (see Fig. 1).

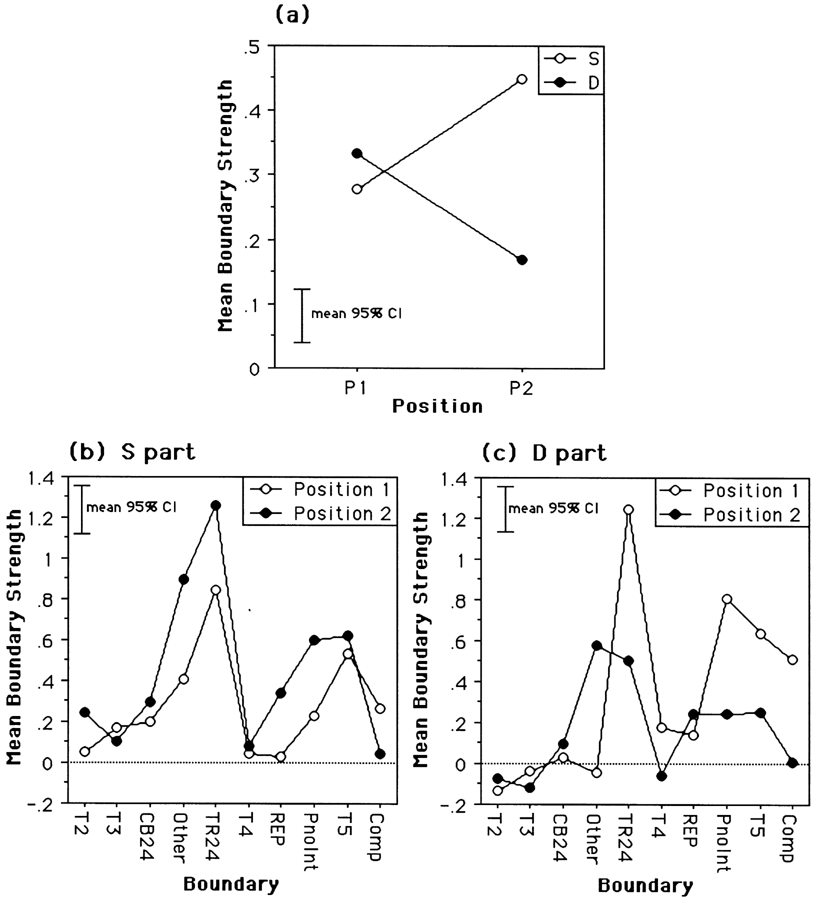

Fig. 7. Mean boundary strength as a function of position within the piece for parts S and D (a) and as a function of boundary for first and second positions in part S (b) and part D (c).

it recognition of the core elements, which occur in nearly identical form in the two parts with only a change in instrumentation.

We derived two measures from the familiarity/resemblance functional objects. In one, the mean rating during the period of the core element was computed. From this value, the mean over the 5-s periods on either side of the core was subtracted, yielding a measure of the emergence of the core element. The hypothesis was that if recognition of the core element occurred, the emergence would be greater in the second than in the first part. In the second measure, the mean of the derivative profiles (not the absolute values this time because we are interested in the direction of change) was taken during the core element and was compared across parts. Again the hypothesis was that if recognition occurred, greater increasing change would occur in the second part.

Both measures were analyzed with a repeated-measures ANOVA on Part (2) × Theme (5) and planned contrasts compared the themes between the two Parts. The analyses were applied to the four performances (DSA1, SDA2, SDU1, DSU2). For the measure of emergence, the core elements of T1 and T5 emerged significantly only in the SDA2 performance, T1: $F(1,84) = 6.1$, $p = .024$, $\varepsilon = 0.758$; T5: $F(1,84) = 11.8$, $p = .0025$, $\varepsilon = 0.758$, and T3 emerged marginally in SDU1, $F(1,156) = 3.6$, $p = .062$, $\varepsilon = 0.909$. None of the other tests were significant. For the mean state change measure, only one test was even marginally significant: T3 in DSA1, $F(1,112) = 4.2$, $p = .053$, $\varepsilon = 0.775$. Therefore, while some core elements appear to be implicitly recognized according to some measures in some of the performances, the recognition is not systematic for all themes across all performances, confirming the general difficulty revealed by Poulin-Charronnat et al. (2004) in explicitly recognizing the thematic materials under instrumentation change.

Summary of Main Findings for Familiarity/Resemblance

• There were global similarities between the mean response profiles for a given version across concerts, despite differences in interpretation, concert hall, cultural milieu and language of the audience, and instruction/training protocol.

• The presence of the computer-processed layer did not influence implicit recognition of musical materials carried by the instruments.

• Participants reacted to local contrasts (as indicated by a sudden dip in judgments when material changed) and to returning materials (as indicated by rises in judgments, at times subsequent to a sudden dip, once material is recognized as similar to those previously presented).

• Sudden changes in the profiles (and peaks in the derivative profiles) occurred most often on section boundaries, although not all boundaries were responded to.

• There was almost no systematic implicit recognition of the thematic core elements played with different instrumentation across the S and D parts.

• There were more prominent *decreases* in the response profiles at section boundaries in the first part of the piece and the first version heard (detection of new materials), whereas there were more prominent *increases* in the profiles at boundaries in the second part and version (recognition of familiar materials).

• There were several effects of the large-scale structure of the piece (part, position, version) and the concert (order) on response profiles: ratings for a given part (S or D) in the version heard second were greater on average than those in the version heard first and were greater on average in second position within the piece than in first position. These effects were stronger for S than for D. For S, the effect of order was stronger than that of position, whereas for D the reverse was true.

• Large-scale context effects were also revealed in the boundary strength measures. There were more significant boundary detections in the D-S version than in the S-D version and more in the S part than in the D part. S boundaries were *more* salient after having heard D, but D's were *less* salient after having heard S.

DISCUSSION

Listeners heard a full piece of music in two versions played by professional musicians in a concert hall while performing a continuous response task. A continuous rating scale probed listeners' evolving sense of the familiarity or resemblance of the musical materials being heard to those heard from the beginning of the piece. This scale seems to capture several levels and time scales of familiarity, resemblance, and recognition. There are reactions to change in material, texture, and sound palette across section boundaries that are related to local differences between what is being heard and what is currently in working memory. These reactions demonstrate the alertness of the listeners to change when it is detectable. The direction of these reactions reflects longer-term memory processes involved in the recognition of returning thematic materials (increased ratings), or completely new and previously unheard materials (decreased ratings). There are more prominent descents in the profile in the first part and the first version heard, indicating detection of new materials. There are also more prominent rises in the profile in the second part and the second version heard, indicating recognition of returning materials. These features of the mean profiles suggest a continual effort on the part of the listeners to assess what they are hearing, which is reassuring.

However, there is an interesting interaction between the reaction to change and recognition that is evident in the mean profiles. At several points, there is a sudden brief descent at a section boundary at which the listeners detected a short-term change in the musical surface, but which is quickly followed by a strong rise as they realized that this material had been heard before. Refer, for example, to Figures 3 and 4. The first occurrence of Other and D10 in SDA2 (second hearing), the second Other (in its D1 computer version in the second part of the piece), and the second piano Interlude in DSA1 and DSU2. There is also a cumulative effect of resemblance across the two parts of the piece and across the two versions heard in the same concert. This accumulation reflects perhaps a more abstract sense of familiarity with the musical style of the composer, with which the majority of the listeners were undoubtedly unfamiliar at the outset. There are thus apparently several kinds and levels of musical similarity that enter into these ratings (see McAdams & Matzkin, 2003, for a review of this issue).

Another factor that arises from these comparisons is that the position effect seems primarily related to previous hearing of the materials in the other part or other version and not much affected by the presence of the computer layer, which always accompanies the second part. The composer was not surprised that the computer layer did not interfere with the perceptual and structural experience of either part. He feels that one is less likely to register the actual, detailed shape of the computer "images" because their ideas are usually spread over a greater time span, and their more subtle details are often obscured by, or blended into, the more assertive instrumental layers. The computer's strongest moments are either as a solo (D10, D1, S7), or as specifically supportive of the instrument layers, as intensifiers towards the end of a section already characterized by the instrumental materials.

Boundary strength measures were derived from the response profiles at the points in time corresponding objectively to analogous section boundaries in both S and D parts of the piece. This use of the derivative of continuous familiarity/resemblance ratings provides a tool for studying the dynamics of perception and recognition. Many of the structural boundaries were reacted to by sudden changes in the profiles that showed up as peaks in the derivative profile. These peaks indicate that listeners were sensitive to certain aspects of the macrostructural organization of the piece as conceived by the composer, in spite of lack of previous experience of many audience members with contemporary music in general or Reynolds music more specifically. They seem to us to be clearly related in many cases to discontinuities in material, texture, tempo, temporal density, dynamics or sound palette or to involve medium- to long-duration pauses (see Figs. 5 and 6, Table 3). Examples include the onset of Other (in S), REP, PnoInt, and the core element from Theme 5. Boundary detection also reflects the return of emblematic material that has a strong structural importance, such as Other and PnoInt in the second part, and the onset of the piano Epilog at the end of the piece. This latter result is remarkable in that these materials are relatively unforceful and undramatic, yet acquire for the listener a certain "weight" due to their recognition. The absence of such changes at other boundaries (notably in T2 and TR13 in D and CB24 and T4 in both S and D) reflects the relative continuity of the musical discourse even though the materials were conceived in a sectional manner in the planning of the composition.

One surprising result is the absence of perceptual boundaries (peaks in the mean derivative profile) at T2 and TR13 in the S part. Both are objectively characterized by changes in instrumentation and texture, and both have noticeable pauses. It is likely that in a segmentation task, a bound-

TABLE 3
Characterization of Changes Across the Sectional Boundaries Indicated in Figure 1

Boundary	S	D
T2	I, MT, P Qualitative discontinuity + pause	I Qualitative continuity
TR13	I, MT, D, P Qualitative discontinuity +pause	— Strong qualitative continuity
CB24	I Qualitative continuity	— Strong qualitative continuity
Other	I, MT, TD, D, P Strong qualitative discontinuity	I, P Weak qualitative discontinuity + pause
TR24	I Weak qualitative discontinuity	I Weak qualitative discontinuity
T4	— Strong qualitative continuity	— Strong qualitative continuity
REP	I, MT, D, P Qualitative discontinuity + pause	MT, D, P Qualitative discontinuity + pause
PnoInt	I, MT, TD, D Qualitative discontinuity	I, MT, TD, D Qualitative discontinuity
T5	I, P Weak qualitative discontinuity + pause	I, P Weak qualitative discontinuity + pause
Comp(D10)	— Qualitative continuity, computer layer fades in slowly	I Weak discontinuity in instrumental part, computer layer fades in slowly
Comp(S7)	— Qualitative continuity	I, MT Weak discontinuity

NOTE—I = Instrumentation, MT = musical texture, TD = temporal density, D = dynamics, P = existence of a pause or silence. Computer section D10 starts at the end of the T5 core element in the first part, and section S7 starts 3.5 measures after the end of the T5 core element in the second part; so the nature of the change depends on the position, not on the part.

ary would be placed at these points (cf. Clarke & Krumhansl, 1990; Deliège, 1989). However, in some way, the pauses (partially filled with resonances from the material preceding it) seem to minimize the discontinuity, and there is a relative rhetorical continuity across the gap. Other cases of interest include Other in the D part, TR24 and T5 in both parts, and D10. These boundaries have only a weak qualitative discontinuity in the materials (Table 3), but are perceived quite strongly as moments of change. They are all quite distinctive in musical character at their onsets, and all except Other have sharp increases in familiarity upon second encounters. This latter result demonstrates the role of dynamic memory processes in macrostructural perception in the absence of strong discontinuities in the musical surface. Taken together, these results suggest that this technique is sensitive to structural properties at different levels of abstraction from the sound events, including qualitative and temporal discontinuities, structural markers such as repetition and return, and rhetorical functions such as continuity, development, and digression (see Lalitte et al., 2004).

These data are very rich and there are many aspects of them that could be addressed further, but we will confine the rest of the discussion to what they indicate about the effect of large-scale form on dynamic perception and memory processes.

Different Memory Coding for S and D Parts

There is a combined effect of order of presentation and position on the response profiles for the S and D parts. For the D part, the second position is rated higher on average than is the first position, in spite of the fact that the former is the first and the latter the second version heard (e.g., SDU1 > DSU2). Further, the second position in the second version heard is rated much higher on average than is the first position in the first version heard (e.g., SDA2 >> DSA1). The position effect is thus stronger than the order effect for D. The inverse is true for S where the order effect is stronger than the position effect. This pattern of results suggests different kinds of memory processing for the two parts, which is most likely related to the very different way the presentation and development of the thematic materials is approached in each. The memory representation of the more diffuse, organic D part is sensitive to shorter term effects, whereas the more didactic and sectional S part allows for longer term effects. We propose that this difference in memory coding and in the interaction of the memory trace for each part with subsequent musical information processing can explain many of the effects of large-scale structure on instantaneous ratings.

Effects of Large-Scale Structure on Perceived Boundary Strength

Large-scale structure had significant effects on local boundary strength, namely, effects of part, position, and order of hearing on the degree to which qualitative discontinuities were detected and reacted to. On average, the boundary strengths were primarily affected by part and by position. Boundaries were stronger in the S part than in the D part, and in the D-S version than in the S-D version. And these two factors interact. Independently of the order in which a version was heard, the S and D parts had equivalent boundary strengths in first position, but these strengths increased for S when preceded by D and decreased for D when preceded by S (Fig. 7a). This asymmetry is parallel to that observed from the functional *F* tests and suggests differences in the memory processing of the two ways of developing the thematic materials in the two parts. This asymmetric relation is further supported by the differences in the influence of each part's memory trace on the perceived strength of section boundaries in the succeeding part in the two versions. The more continuous flow of D in first position has less effect on S in second position than the clear sectional presentation of S in first position has on D in second position. The S part has more effect on itself across the two versions heard, which suggests a clearer mnemonic encoding and persistence of the materials in their S realization than in their D realization.

However, these results do depend on the boundary being tested and are not equivalent for all boundaries. There is no effect of position or part for T2, TR13, and CB24 since no boundaries were detected at the beginnings of these sections anyway. For most of the other boundaries tested, there is a pattern of increase or equivalence of strength from first to second position for S and a decrease or equivalence for D (TR24, T4, REP, PnoInt, and T5). There are two exceptions to this rule: Other always increases from first to second position, being not so noticeable and unrecognized in first position, and clearly recognized and noticeable because of this recognition in second position; and the Computer solo, which always decreases from first to second position demonstrating a stronger effect of D10 (the bridge between the two parts) than of S7 (the coda leading to the piano Epilog).

Taken together with the functional *F* tests, these results on boundary strength raise the question of why there are more and stronger boundaries in the D-S than in the S-D versions. One would expect the more didactic, landmarked presentation of S to clarify the material-based structure to a greater extent than would the more diffuse character of D. As such, a more precise and enduring memory trace would be established for S that could then interact with incoming musical information over longer time spans. This idea has implications for the formation of memory representations and the priming force of previously constructed representations as a function of their nature (D vs. S) on the rest of the work (cf. Vieillard, Bigand, Madurell, & McAdams, 2005, for a similar discussion of priming in electroacoustic transformations of musical materials in *The Angel of Death*).

One explanation is perhaps because listeners do not expect boundaries to come as fast as they do in the first part of the S-D experience or have not acquired enough experience with the style yet. They therefore give less weight to the idea of a change in S-D than they do in D-S. However, similar results occur for S-D when heard second in the concert, in which case listeners would have acquired such stylistic experience. In D-S, change is a more unusual and decisive event in the second (S) part that clarifies and renders more explicit the materials presented in a more fluid form in the first (D) part. The brain is always trying to find order in perceptual stimuli (Michon, 1978). If a listener experiences music that is ambiguous (D) for a period of time, then when perceivable order is introduced, it will most likely increase attentive behavior, which can benefit from the even diffuse previous presentation of the thematic materials in processing new materials and structure, thus consolidating them in memory. When S comes first, a clear representation of the materials accumulated in memory is subsequently perturbed by a more varied, interacting, and less clear-cut presentation of the materials. The continuous discourse in the D part does not lend itself to segmentation, categorization, and recognition. As such, the sense of familiarity/resemblance is affected by the memory trace of the first part to a lesser extent than occurs for the reverse order of the parts.

Memory Dynamics

These results raise issues concerning memory dynamics in sequentially structured activities. Dowling, Tillmann, and Ayers (2001) found results for music that differ from those found for reading by Sachs (1967). The latter study showed less memory for surface detail but retention of gist and meaning (although, to the contrary, Goldinger, 1996; Goldinger, Kleider, & Shelley, 1999, showed retention of surface detail). Goldinger (1996) has noted that memory for surface detail was more often revealed by implicit than explicit measures, and these are differentially affected by the passage of time. Surface details fade rapidly from explicit memory (involving the recall of items, perhaps under cognitive control) but persist in implicit memory (which affects behavior automatically without explicit recall). For example, Dowling et al. (1995) found that contour encoding was controlled, but pitch interval encoding was automatic: discrimination of target and new, unrelated melodies was subject to interference from concurrent tasks, whereas that of targets and similar-contour melodies was not. They claimed that the recognition improvement in the target/similar-contour task resulted because while listening continues, the processing of already-heard material proceeds automatically, increasing precision of memory representation of what was already heard. So the dynamics of memory are affected by the degree of structure and meaningfulness (perhaps musical coherence) of intervening materials.

This theoretical framework seems coherent with the differences observed in the familiarity/resemblance profiles and in implicit boundary detection in the S-D and D-S versions of the piece. It suggests that the structural nature of S facilitates encoding in memory to a greater extent than does that of D. It remains to be seen what the importance of the meaning and musical structure of intervening materials is. Is it only that they increase the engagement of the listener or does some association of original and later occurring patterns play a role? The flux of perception and memory creates a flux of experience depending on what the listener is paying attention to and/or reflecting upon while listening, as well as the time elapsed since hearing a given pattern initially. What changes continually during listening is what one can remember having heard. The present results extend those of Dowling and colleagues in demonstrating effects on implicit recognition over time spans of tens of minutes between parts and more than an hour between versions.

Gerrig and McKoon (2001) propose the notion of resonance, a memory process that ensures experiential continuity across temporally discontinuous episodes. Resonance is a fast, passive, and easy process by which cues in working memory interact in parallel with, and allow access to, any of the information in long-term memory. Different information in long-term memory is evoked to differing degrees according to the strength of association with cues in working memory. Accessibility changes over time continuously as successive cues enter and leave working memory. One is reminded of many cinematic examples here with the filmmaker weaving together several, temporally interspersed viewpoints over a couple of hours. Gerrig and McKoon propose that the accessibility of the discourse entities depends on this resonance process at different points in stories with episodic structure (a sort of waxing and waning of entities over the course of a narrative). This memory-based approach to text processing demonstrates that significant aspects of comprehension do not require readers to engage in special, goal-directed behaviors. They provide the foundation for inferences about common ground and characters' perspectives (Gerrig & McKoon, 1998). Resonance binds episodes together by making prior, related experiences accessible in memory even as the current episode is evolving. This may have import for listening to musical pieces with structures such as the S part of *The Angel of Death,* allowing for both contrast and experiential continuity as threads of the "personages" of the story (the thematic materials) are re-encountered at later points of the story, allowing for the interaction between perception and memory proposed by Tulving (1983). The differences in memorability of the structures themselves may indeed strongly affect the time course of these resonances, giving rise to the asymmetric effects of large-scale form demonstrated in our results.

Emotional Force Scale

RESULTS

A different analysis approach was used with the emotional force profiles, because missing data occurred during periods in which some listeners moved the slider into the "Don't know" zone. The current state of tools available for functional data analysis does not allow for automatically fitting functions to data with missing values, and it was not feasible to do it by hand with the amount of data to be analyzed here. We therefore simply used the scaled raw data for the statistical analyses. Any statistics reported do not include missing data. *F* functions were computed to compare groups of profiles that were registered to the same timeline. It should be understood that the number of observations included in the computation varies over time (mean number missing = 8.4, 1.7, 1.2, 1.4; max number missing = 34, 10, 10, 10, for DSA1, SDA2, SDU1, DSU2, respectively). For establishing decision criteria, the mean number of observations over the duration of the profile was used. The mean scaled emotional force profiles are shown for the two versions of the piece in the two concerts in the upper panels of Figures 8 (S-D version) and 9 (D-S version).

There is less variation in the mean emotional force profiles than in the familiarity/resemblance profiles. The mean profiles vary over about half the full scale, centered near the middle. Given that the raw data were scaled to the same full range [0–1], this limited range of the mean profiles indicates that there is a great deal of uncorrelated individual variability between listeners, and that only the moments when there is a relative synchrony in their movements will be evident in the mean profiles. There is a strong shaping of the mean emotional force profiles with peaks on the high side evoked, in rough order of mean peak force, by the computer solos (the D10 bridge and the S7 coda), the TR24 region in both parts, the CB24 region in the second part, and with dips on the low side evoked by Other, PnoInt, the T5 core element (just before the computer enters), and the Epilog. Note the similar aesthetic impact intended by the composer for PnoInt, T5 core, and Epilog (see Reynolds, 2004).

The mean profiles can be considered as a hierarchical series of nested arches, the low points of which often, but not always, correspond to section boundaries (e.g., the beginning of Other, TR24, T4, T5, and D10/S7). An example of the hierarchical nesting structure can be seen in Figure 9 for the DSU2 mean profile (thick line). A series of four smaller arches creates a more global arch ending at the beginning of TR24. The nested arches include T1 to one third of T2, the latter two thirds of T2 to one-third of CB24, the latter two thirds of CB24, and Other. Some of them have even finer arch structure in them. The next group of four or five arches ends at the beginning of D10, and so on. Seen in this way, the four performances appear to have five to six larger sections with an increasing then decreasing emotional shape, within which are nested one to five smaller arches. It is the characterizing of this shape of emotional experience that interested the composer at the outset of the project. Note that while the thematic elements play a strong role in the evolution of the sense of familiarity and resemblance, the strong sections in an emotional sense

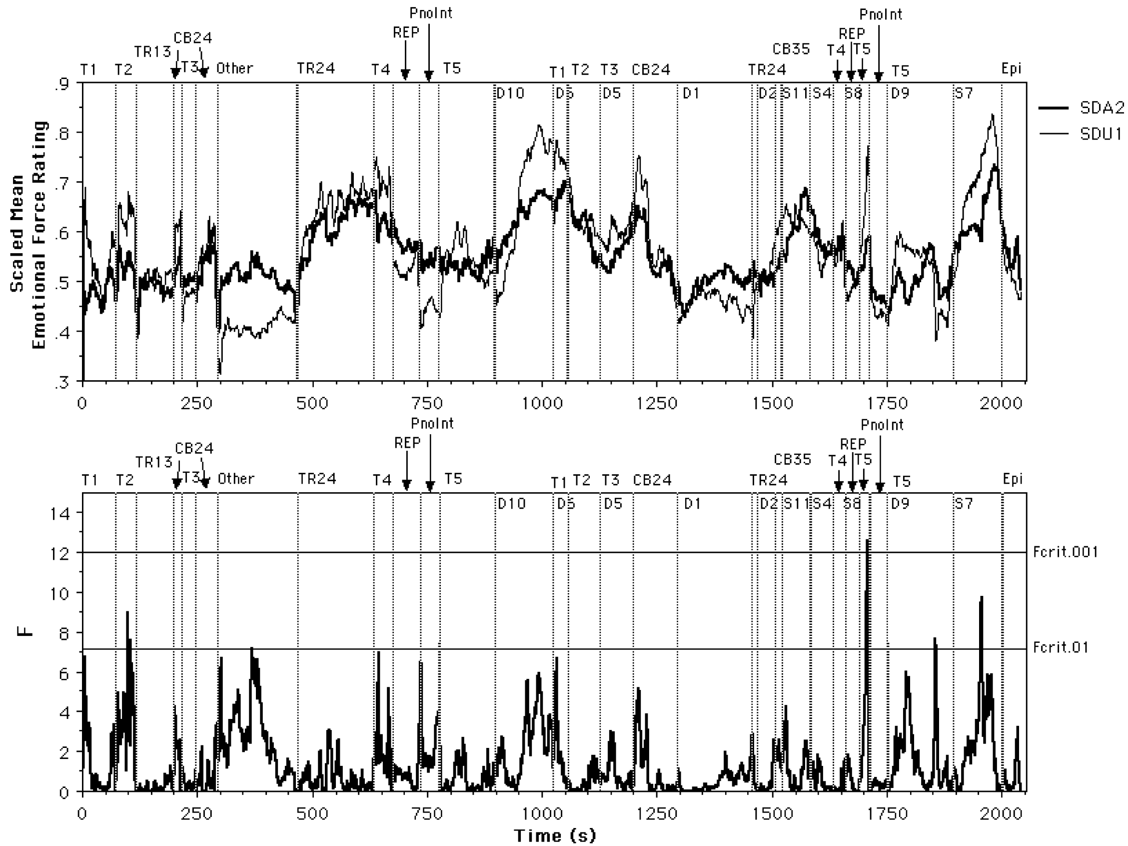

Fig. 8. Mean scaled Emotional Force profiles for the S-D version and the functional *F* test representing their statistical comparison. (see Fig. 3 caption)

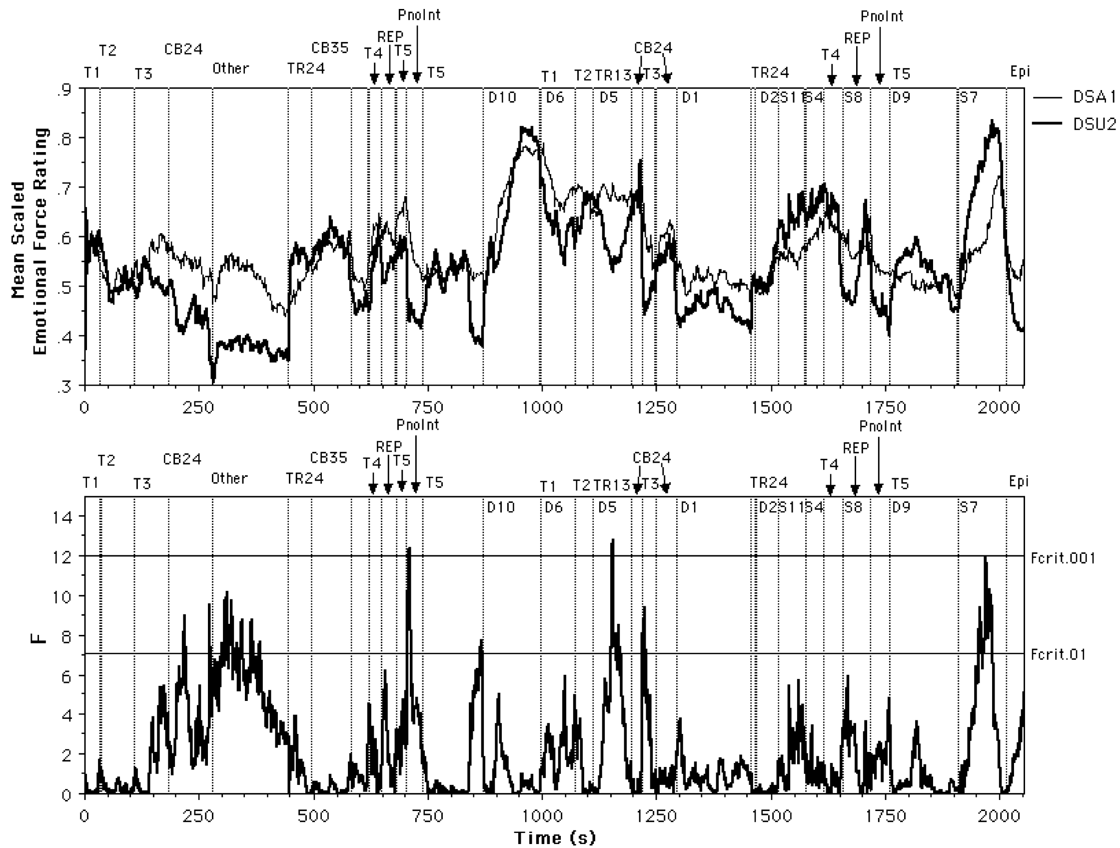

Fig. 9. Mean scaled Emotional Force profiles for the D-S version and the functional *F* test representing their statistical comparison. (see Fig. 3 caption)

(higher mean emotional force ratings) are the derived sections: Transitions, Combinations, and solo computer sections (D10 and S7).

Subsequent analyses will focus on comparisons between the concerts (different orders of hearing the two versions) and between given parts (S and D) in first and second positions within the piece.

Comparisons between Concerts

Globally the mean emotional force profiles are very close between the two performances by different ensembles, with only 99 s of significant divergence for D-S and 9 s of divergence for S-D at the $p = .01$ criterion (2.5 s and 1 s, respectively, at the $p = .001$ criterion). There are wider and more abrupt swings in the mean profiles for the La Jolla performances than for the Paris performances, suggesting greater response synchrony in the former group. This could be due to a greater responsiveness of the listeners or to greater evoked emotional variation in the SONOR performance. Some of the differences that characterize the two interpretations are a more subtle, refined, and blended sound for Court Circuit in Paris, and a more dynamic, contrasted sound with greater bringing out of inner voices for SONOR in La Jolla. Also the computer layer was much more prominent in the La Jolla performance, because the hall was larger and the space allowed for higher levels.

In the S-D version (Fig. 8), the profiles are very close and only significantly diverge momentarily. These divergences are due to a stronger presence and more dramatic use of brass and percussion (in T2 near the beginning, and just preceding PnoInt in the second half), to a slower more mellifluous rendition of the T5 core element, and to a more prominent computer layer in S7, all in the La Jolla performance.

In the D-S version (Fig. 9), the profiles for both concerts start at the same emotional force level. They diverge shortly after the entry of T3 with a more restrained tension in DSA1 compared to a more straightforward rendition in DSU2. Then DSU2 descends into a particularly tranquil and floating Other. The profiles rejoin at the beginning of TR24. There is a rise in the T5 section preceding PnoInt in DSA1 due to a greater crescendo and dissonance, and a quicker and greater descent at the entry of PnoInt in DSU2 due to a longer silence preceding a softer rendition of this interlude. A descent at the end of T5 in DSU2 gives another significant peak in the F function, again due to a longer silence that demarcates the beginning of the T5 core element. There are momentary divergences in TR13 and T3 with DSU2 moving toward lower emotional force levels, which seem related to the presence of the D5 computer section in TR13, as well as a silence preceding a very soft, fluid rendition of T3 with a lot of sustain pedal. The profiles then stay fairly close until S7 where a greater rise in DSU2 is due to the higher levels of the computer layer at the La Jolla concert.

Comparisons of the Same Part across the Two Versions

There is a higher correlation between the profiles of a given part in the two positions within the piece (e.g., D in first position in D-S and in second position in S-D) for the La Jolla audience than for the Paris audience (La Jolla: $r^2 = .54$ for S and .34 for D; Paris: $r^2 = .03$ for S and .08 for D). As Schubert (2000) has remarked, correlations between time series should be interpreted with caution: they tend to be inflated because they do not take into account serial correlation, that is, adjacent samples are not independent. The S part in the Paris concert (Fig. 10, left panels) was rated as having greater emotional force in D-S than in S-D over T1, T2, TR13, and the beginning of CB24. Because the computer layer is not very prominent in this region and blends into the instrumental layers, and because the performances seem fairly similar over this part of the piece, the higher mean profile in DSA1 may be an emotional holdover from the strong peak in the D10 computer solo immediately preceding the S part in this version. The difference between the versions is inverted thereafter but only reaches significance momentarily at the $p = .01$ level near the beginning and middle of TR24. There seems to be a stronger relation to the piece's structure in the S-D version with sudden dips in the mean profile at section boundaries. For the S part in the La Jolla concert (Fig. 10, right panels), there is a quite different pattern. S-D has higher mean emotional force ratings over TR24, T4, REP, and T5 (without the computer layer). It is not clear from listening to the recordings of the two performances to what this difference might be attributed. In this region, and in both parts and both versions, the version without the computer layer has the higher mean emotional force. In the La Jolla performance, there is a prominent emotional peak at the end of REP in D-S that is clearly due to accelerating density and level in the S8 computer section.

For the D part in the Paris concert (Fig. 11, left panels), the profiles are close but uncorrelated, wavering in a small range in the middle of the scale until T4-REP-T5-PnoInt where D-S is higher (again without the computer layer). For the D part in the La Jolla concert (Fig. 11, right panels), there are regions of strong divergence at the beginning of CB24 with S-D being higher due to the end of the D5 computer section. Another divergence occurs at the beginning of TR24 with D-S being higher apparently because of a particularly dramatic percussion entry (a tam-tam shriek produced by scraping the tam-tam surface with the end of a wooden mallet). In the REP section, S-D descends abruptly due to a longer silence preceding REP, and then abruptly rises again during the crescendo and increasing density of the

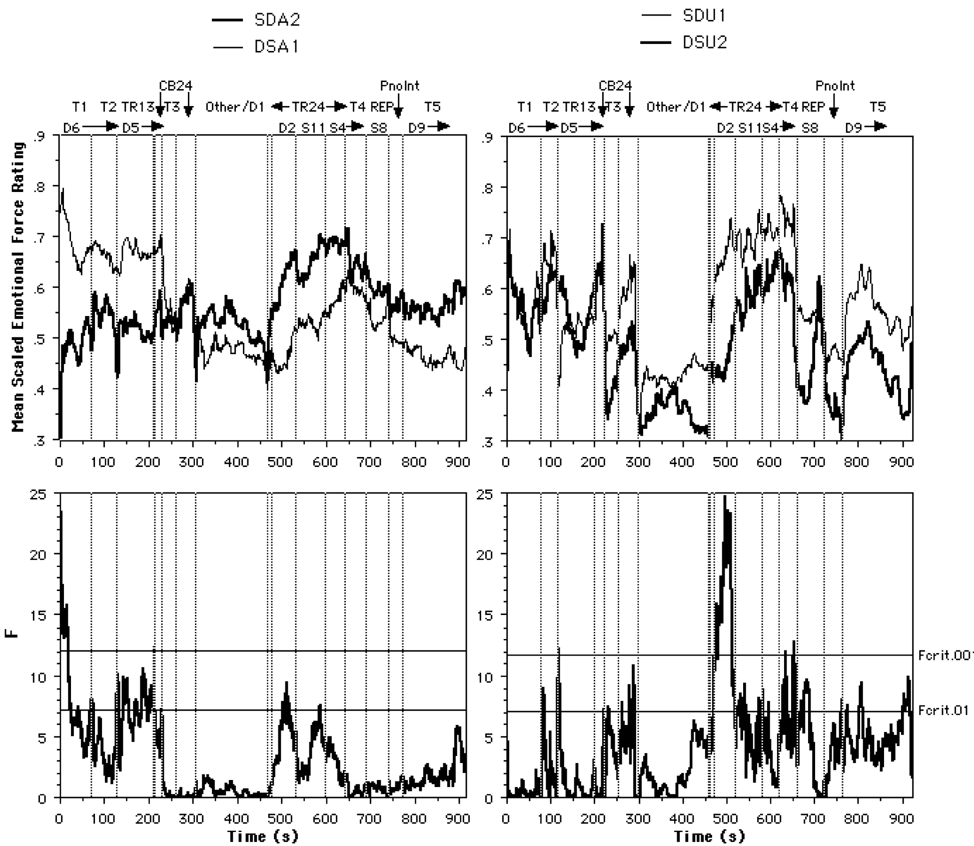

Fig. 10. Mean scaled Emotional Force profiles for the S parts of S-D and D-S versions and the functional *F* Mean scaled Emotional Force profiles for the D parts of S-D and D-S versions and the functional *F* test representing their statistical comparison. Paris concert = left panels, La Jolla concert = right panels.

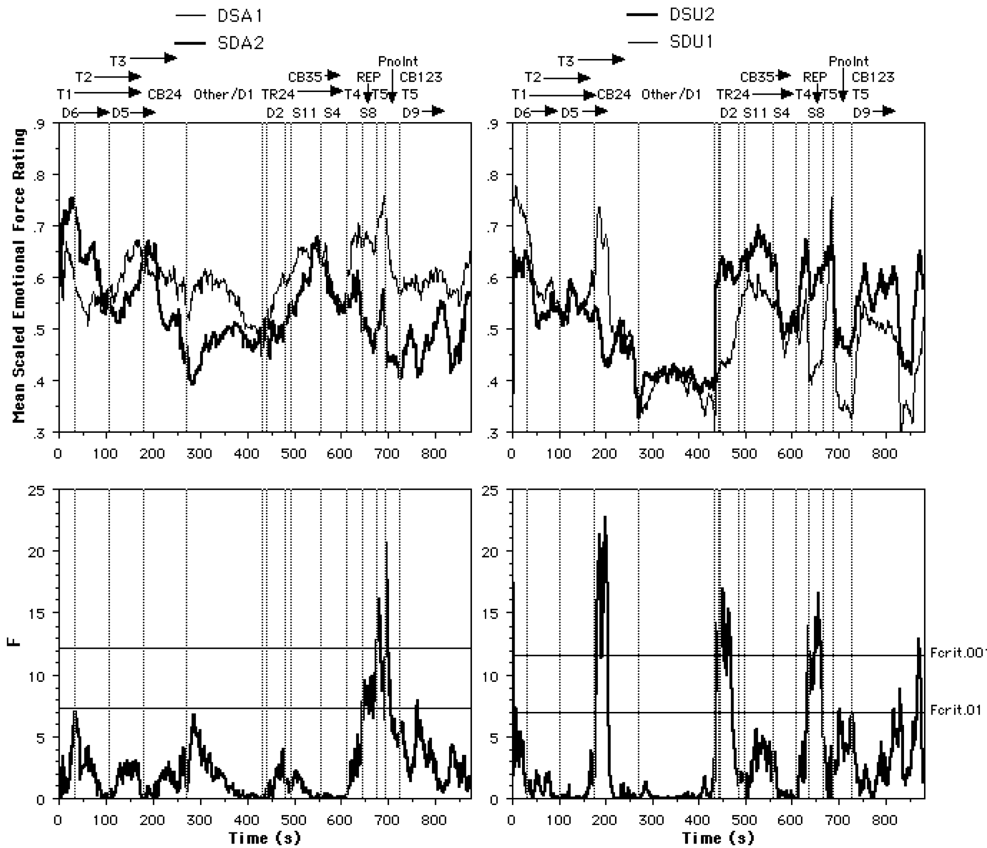

Fig. 11. Mean scaled Emotional Force profiles for the D parts of S-D and D-S versions and the functional *F* test representing their statistical comparison. Paris concert = left panels, La Jolla concert = right panels.

S8 computer section. This creates a strong subsequent contrast with PnoInt, hence the abrupt redescent of the profile. This latter effect parallels that in the S part for the La Jolla concert (Fig. 10) mentioned previously. At the end of T5, the onset of D10 in DSU2 apparently creates a stronger emotional response than does the entry of S7 in SDU1.

If we examine the differences between mean profiles for the versions for each concert, we can glean something about the effect of position of the part in the piece and the order in which the version was heard in the concert. For the S part, the S-D version has higher mean emotional force ratings than the D-S version for both concerts, whereas for the D part, the D-S version has higher ratings than the S-D version. If compared to a theoretical difference of zero with a single-sample *t* test, these differences are all highly significant ($p < .0001$).[4] The means are shown in Table 4. From these values we can see that a part in first position is always rated higher than when it is in second position, and further, this difference is enhanced when the first position is in the first version heard compared to when it is in the second version heard. There is thus a dominance of the position over the order, but both suggest that emotional force ratings diminish with repetition.

Summary of Main Findings for Emotional Force

• There is a close resemblance of the emotional force profiles across performances for a given version.

• The strongest emotional peaks occur in the computer solos and regions derived from Theme 2 and 4 materials: TR24 and CB24.

TABLE 4

Means of the Differences between Positions for a Given Part

Part	Comparison	Positions	Orders	Mean	*t* (*df*/*df′*)
S	SDA2 – DSA1	1 vs. 2	2 vs. 1	0.013	5.0 (1831/15.2)
S	SDU1 – DSU2	1 vs. 2	1 vs. 2	0.081	47.8 (1849/15.4)
D	DSA1 – SDA2	1 vs. 2	1 vs. 2	0.065	34.3 (1743/14.5)
D	DSU2 – SDU1	1 vs. 2	2 vs. 1	0.034	16.1 (1769/14.7)

NOTE—*df′* are the degrees of freedom corrected for serial correlation.

4. This significance criterion should be interpreted with caution due to the serial correlation inherent in time series data. If the autocorrelation functions are standardized to a maximum of 1, the roughly triangular form falls below .75 at between 30 s and 60 s. We could consider that samples separated by roughly 60 s are independent. If the degrees of freedom are reduced accordingly, they are 14–15 for the four versions of the piece. With these degrees of freedom, all of the *t* tests are still significant at $p < .0005$.

• The collective emotional force profiles have a form that can be likened to a hierarchically nested series of arch forms.

• Emotional force decreases globally from first to second part and from first to second version heard.

DISCUSSION

Continuous ratings of emotional force were measured for the two versions of *The Angel of Death* in two premiere concerts. According to Schubert (2001), this scale probably taps primarily into the arousal component of emotional reaction. This is confirmed in part by the modeling of the acoustic origins of emotional force reactions in Dubnov, McAdams, and Reynolds (in press), as a combination of information rate, energy, and spectral centroid (faster rate, higher energy, higher spectral centroid = higher emotional force). It should be noted, however, that a simple equation between emotional force and energy, as proposed by Schubert (1999, 2001) for example, is too simplistic, because there are clear moments in the piece where energy is low and emotional force is high and vice versa. It is the combination of parameters, as well as the more global trajectory of the musical discourse, that appear to determine the emotional response. Indeed for tonal music (Mozart), Krumhansl (1996) found that tension ratings were pretty much the same after taking out changes in loudness and/or tempo. That would mean that harmonic structure was quite important. Of course in contemporary music, loudness and tempo come to play a more important role in emotional experience, in the absence of harmonic structures that are overlearned as in tonal music.

In spite of the fact that a majority of the listeners, and most particularly the nonmusicians in the audience, were unfamiliar with contemporary music in general and Reynolds's music in particular, the music clearly affected the participants emotionally as evidenced by individual emotional force profiles with strong variation over the course of the piece. Further, there is a strong relation between the mean profile (representing collective response) and the structure of the piece, although the structure revealed by the emotional force data diverges in interesting ways from that revealed by the familiarity/resemblance data. One of the striking features of the mean profile is its apparent organization in a series of hierarchically nested arch-like forms. The emotional force profiles reveal the relative importance of computer solos (D10 and S7) and developmental sections (particularly TR24 and CB24) for high emotional impact, as well as the reflective periods of emotional repose on the low end of the scale in Other, the piano interlude, T5's core element, and the final dénouement with the Epilog. The arch forms recall performance phrasing profiles (Penel, 2000;

Todd, 1995). However, the latter operate at the phrase level, whereas suprasectional time spans are involved in the present study.

The nested arch structure found in the emotional force data is also analogous to hierarchies of structure and affect (often discussed in relation to musical tension and release cycles) found in tonal music as described by Lerdahl in his Tonal Pitch Space theory (Lerdahl, 2001). The presence of natural modulations in emotional response suggests that there are similar dynamics of experience in nontonal contemporary and traditional tonal music, although they are probably engendered by different structural properties in the music. More broadly, there seem to be invariants in the nature of human affective response in general that are tapped into by these different genres of art performance. Until recently, music cognition research focused on Western tonal music, but increasing attention is being devoted to diverse musical cultures (e.g., Ayari & McAdams, 2003; Balkwill & Thompson, 1999). It is interesting to see similarities across performance genres such as those revealed in this investigation. Through this research we may come to understand more about the variants and invariants of human emotion and experience.

The composer finds the fact that the emotional force tends not to derive from thematic elements per se entirely reasonable. He feels that it is often the case that it is not the posited element (the fact, the identity, the motive, the character, the theme) that moves us, but rather what happens to the things we have recognized or internalized: transformation not identification leads to emotion. He said that in composing the piece, he was of course thinking about nested emotional hierarchies (though perhaps not as a conscious goal): for example, of T1 being assertive and linear, followed by a T2 which disrupts while maintaining the energy level, the transiting from a state of disruption by milder, continuous alternations (modified interruptions) in TR13 that lead to a gentle, whirring close in T3. The idea of searching for arches of emotive response could be a fruitful way of looking at the design and effect of musical experience. Future work could look at how such a model of the judicious shaping of the emotional context compares across musical styles and periods, particularly as concerns temporal proportions among emotional arch-shaped sections.

The general emotional shape is fairly similar across interpretations (perhaps due to the composer's presence at all of the rehearsals) with some notable differences that for the most part can be attributed to the conductor's interpretation (the orchestral rendering styles of Valade and Sollberger are quite different in this respect). In spite of variation in a number of factors, including performance venue, performers, audience members, conductors, intervening musical pieces, and cultural context (Paris, France vs. San Diego, California), the emotional "fingerprint" of the composition was preserved.

However, what is striking here is that the differences are a modulation of a large-scale emotional form that is closely tied to the musical structure. The La Jolla performance did seem more "emotional" to those of us who were present at both concerts. Further, it probably had a better overall "sound" because the Mandeville Auditorium is larger than the Grande Salle. It could be that the increased physical/visual perspective gave listeners a more "secure" basis upon which to judge, combined with the fact that the seating was more spacious and listeners in a less tight physical proximity to each other in La Jolla. If the audience members' movement was restricted in the Grande Salle, that might have caused a reduction in the emotion experienced. Research has shown that when emotional expression is inhibited, the experience of emotion and related physiological changes will be dampened (Hatfield et al., 1994). The kinds of expressive movements that audience members might make include sways, postural adjustments, and other such movements. Another potential source of the difference between concerts may be that the visual aspect contributed to an increase in emotional experience. Vines et al. (2003) have found that being able to see a performer can augment emotional experience (as measured by musical tension ratings) during a performance. However, the Mandeville audience being larger, and listeners farther away from the stage, would suggest the inverse of what was observed in the data. Given that the difference in rated emotional force was not general across the piece, but reflected more in the abruptness of changes in the mean profile, it seems more likely that differences are due to the interpretation. Studying the derivatives of these profiles may provide interesting information along these lines if the problem of fitting functional data objects to profiles with missing data can be solved in the future.

One of the more striking features of these data concerns the effects of previous exposure to the materials (in a preceding part in the piece or a preceding version in the concert). In line with Meyer (1956), one would expect expectation and surprise to play a strong role in emotional response, which could develop with repeated exposure to the musical grammar and style. The emotional force depends to some extent on the position of the parts and less on the order of hearing of the versions, but both effects tend in the direction that repetition diminishes emotional force slightly. As such, one wonders if the change is due to something related to satiety. While there is little exploration of the role of satiety in emotional response, theorizing on its role in preference and liking has been done. According to Berlyne (1970), the evolution of preference with exposure depends on the initial level of activation potential of the stimulation (most likely associated with arousal in the present case) and in particular on the initial perceived complexity (contemporary music) and novelty (premier concert). The basic idea is that appreciation diminishes with a high level of familiarity. This may be related to the notion of psychological complexity (Dember & Earl, 1957). Individuals prefer stimuli with a complexity level corresponding to their optimal level of psychological complexity, which depends on and increases with experience (and learning). Although such considerations may be relevant over many hearings, it is hard to imagine that they operate over the two parts of a piece or across two versions in the same concert. Given the high complexity of the music and the attentional and vigilance demands of the continuous rating task (over two periods of 35 min), it may be that the experimental situation played a role in the relative decrease in emotional force with repeated exposure. This interpretation fits with Schubert's (1996) suggestion that habituation could lead to a reduction in emotional force due to a gradual increase in activation threshold of "emotion nodes" in a network of the general type proposed by Martindale (1984) to explain the appreciation of negative emotions in aesthetic contexts. The idea that repetition diminishes emotional force is perhaps not so problematic at the "first-stage" or early encounter stage that is under consideration here. One would hope or expect, however, that, over time, as one began to integrate all of the aspects of a work and especially in a compelling performance, that one would recover and even enlarge upon the original novelty-influenced emotional force experience. Experiments are planned to look at the effects of repeated exposure over several listenings across several days on continuous ratings, which are particularly cogent for a musical style with the inherent complexity of *The Angel of Death*.

General Discussion

The Angel of Death embodies a musical form with a large number of material and structural properties that provide for any of a number of dimensions of musical experience. One of the main interests of the composer within the Angel Project was to explore what he imagined people's experience of musical form to be in response to the piece. That is, how do a listener's impressions evolve over time and to what extent does the architecture of the piece leave her or him changed? The idea of evolving impressions is not to be taken as equivalent to "noticing" the formal structure, but rather to living through it in time. Behind this quest are the questions of what properties of musical form maximize a listener's implicit grasp of and response to the shape of their experience and the ideal levels of ambiguity that maintain interest over several tens of minutes.

DIFFERENT ASPECTS OF THE SHAPE OF MUSICAL EXPERIENCE

The two aspects of musical experience that we have attempted to probe in continuous fashion through time are related to memory and emotional processes. The large-scale formal relations that can be compared in the two versions of the piece had quite different effects on the responses along these two dimensions, demonstrating listeners' sensitivity in instantaneous response to such formal relations along the scales of familiarity/resemblance and emotional force. Changes in the familiarity ratings were related to section boundaries in the piece and changes in thematic material. The emotional force judgments were not bound closely to the prescribed sections, but a consistent global contour emerged in response to the music, which was consistent across performances, musical ensembles, conductors, and venue. The piece itself would seem to have an emotional "fingerprint" in a sense similar to the way Madsen discusses responses to music (cited in Fredrickson, 1997). This basic underlying emotional shape is expressed with variations that depend on the variables mentioned above: venue, performing ensemble, soloist, conductor, and audience culture. However, in spite of there being variation across participants' judgments, a similar underlying emotional force profile emerged for the two performances.

Familiarity profiles show a much greater effect of previous exposure (effects of position and order) that interact strongly with the nature of the materials being rated (S or D part), suggesting very different memory dynamics for the two kinds of realization. The finding concerning the differences in mean familiarity profiles and boundary strengths between S and D parts and between S-D and D-S versions should be considered in relation to the general preference of listeners in the questionnaire, and of the composer, for the D-S version. What in the implicit perceptual structural analysis of the piece might have contributed to such a preference? Reynolds (2004) initially thought that there would be advantages for the listener in S-D due to being able to hear the materials clearly at the outset in S and to the fact that D would be more transparent because of the distribution of materials between piano and orchestra, making it more amenable to work with the computer layer in second position. As such, he predicted that S-D would be the more satisfying version. But the musical reality was that he found the "oceanic darkness" of the D part to provide a more gripping opening and the TR24 section played by the piano in D to be more persuasive in the first part than was the orchestral analogue in S. So the logical and strategic merits of S-D were swamped in his estimation by more persistent emotional tugging of D-S and the more satisfying arrangement of materials moving toward clarity of structure rather than away from clarity. This structural clarity was clearly demonstrated by the differential effect of large-scale form on the S and D parts with D-S having the strongest perceived boundaries, particularly in the latter S part. So the move from diffuse structural fluidity to clear structural perception may have an inherently more intellectually satisfying "form" in terms of its dynamic contribution to the overall experience.

One question that comes to mind in considering the many different dimensions of psychological experience in listening to music is the relation between them over time. We have measured two aspects. Consider the extent to which emotional force depends on a surprise effect (decreasing familiarity or sudden contrast). If emotional force increased when familiarity decreased, this would support the classic hypothesis of Meyer (1956), according to which the violation of musical expectations plays an important role in the formation of emotions. Accordingly, one would expect a strong negative correlation between the two. In fact, the coefficients of determination (r^2) between the two measures are actually fairly weak, if we only look at parts:

S part: DSA1 = 47%, SDA2 = 2%, SDU1 = 0%, DSU2 = 30%
D part: DSA1 = 4%, SDA2 = 9%, SDU1 = 23%, DSU2 = 1%.

And they only increase slightly in some cases if one looks at smaller sections.

Another approach to considering the multidimensionality of musical experience is to examine the correspondence between the subjective structure of the piece as revealed by implicit boundaries in the familiarity/resemblance profiles and by the low points in the emotional force profiles that are the end points of the archlike forms. The "subjective" musical structures inferred from the mean profiles derived from the two rating scales correspond at times and are different at other times (i.e., highs and lows of emotional force and familiarity line up at times and don't at others). So generally the notion that expectation and surprise affect emotional response as proposed by Meyer (1956) does not seem to play out very strongly in this piece. The sensing of the two dimensions of experience simultaneously in musical listening raises a number of interesting questions concerning musical form and the mental structuring of time that this kind of approach may allow us to investigate more systematically in the future. It is interesting to speculate on the feeling created in a listener when these different aspects of temporal experience converge and diverge over time. They may engender attentional trajectories and expectations that only partially covary and that can tug and push in different directions at the same time as one listens to the music.

METHODOLOGICAL ISSUES

The use of continuous real-time rating techniques in concert listening has both advantages and disadvantages that should be taken into consideration. The advantage of real concert listening is the immersion of the listener among other listeners in a collective cultural experience, with the presence of live musicians with the visual component that watching them includes. It may be supposed that such a situation would facilitate emotional communication and in addition, the listening conditions are optimal. Among the disadvantages one might cite the potential distractibility of listeners caused by the activity on the stage and neighboring listeners. Also the task itself requires that the listener adopt a self-monitoring orientation that may not always be compatible with a free listening experience. That being said, several listeners indicated spontaneously in the questionnaires that the focused task enhanced the quality of the listening experience, because it drew them into the piece more. This recalls the finding of Madsen and Coggiola (2001), who found enhanced appreciation in a continuous response task. One difficulty that listeners using the familiarity scale reported was to inhibit previous knowledge related to other pieces in their ratings. It is clear from the order-of-listening effect that on a second hearing, familiarity ratings were higher on average. As mentioned in the discussion on the familiarity data, this task probably involves both local aspects related to the detection of change (local familiarity or resemblance), as well as more global aspects related to recognition of materials, but also a growing familiarity with the musical style.

The mean emotional force data were fairly close to the middle of the scale, even though the data were individually scaled to occupy the same full range [0–1]. This fact points to the necessity to develop subject classification techniques on functional data that allow us to look at differences between classes of people with similar emotional force profiles. We could perhaps then characterize what aspects of the piece each class was responding to.

In this exploratory study, we have moved along the continuum of experimental design towards the extreme of ecological validity. This is the first study ever to explore the real-time experience of live concert performances. For the composer's artistic integrity and due to practical inevitabilities, the experiment was not as well controlled as a laboratory study. There were potential confounding variables, such as language of the instruction set and differences in the intervening music between halves of the piece that were allowed in order to achieve maximal ecological validity. The fact that the composer worked closely with the conductors and ensembles as they rehearsed may have created greater consistency across performances. We are in the process of analyzing experiments that used recorded versions of Reynolds's piece, presented in controlled listening situations. Those studies will complement the research presented here and will be used to evaluate the robustness of the data collected in a live concert setting. Further, due to organizational constraints beyond our control, we did not have full data sets for both orders of presentation at both concert venues, that is, with audiences of a similar cultural milieu, with the same musicians and in the same concert hall. We thus have a number of confounds between order of presentation and venue that will need to be teased apart in future experiments with recorded versions of the concerts. Another confound is the coupling of position of the part and the presence of the computer layer, which always accompanies the second part. Work in progress will deconstruct this confound by using the same recorded sound signal for a given part in both positions without the computer layer, in order to be able to evaluate the effect of its presence.

A study of participants' musical expertise and listening habits is needed to better understand the contribution of implicit knowledge to familiarity and emotional force ratings separately. Contemporary music may well evoke complex emotional responses that are not completely similar to the most "basic" emotions (melancholy/sadness, elation/happiness . . .) that result from listening to tonal pieces. Although for practical reasons related to the number of simultaneously operating response devices, the current study sought a simple unidimensional scale, it would behoove us to explore the kinds of emotional responses that people have to contemporary music of different styles. It would also be of importance to study the relationship between greater implicit knowledge of the musical style and emotional response. Repeated exposure to a given piece of contemporary music would probably decrease the perceived level of psychological complexity in listeners, a greater appreciation of this style (influence of familiarity on musical preference), and consequently a more contrasted emotional force profile may then result. We have planned experiments to follow up on this aspect both with *The Angel of Death* and other pieces of contemporary, as well as romantic and classical, music.

In closing, it seems that, given the really formidable complexity of the design and the length of this piece, as well as the fact that listeners must realistically be assumed to be very inexpert in relation to the composer's musical style, any effort to reach for too much subtlety in interpreting the deep causes of variability would be misplaced at this stage in the development of these techniques. The important issues are the larger points: plausible familiarity results, and the sensitivity of listeners to boundary points, the effects of large-scale form and the nature of musical materials on memory dynamics, the relation of emotional force profiles to the musical structure, and the differing aspects of the shape of musical experience that are revealed by the two dimensions measured. The questionnaires show that for most, but not all of the listeners, the rating task was relatively noninvasive vis-à-vis the music-listening experience, and indeed some seemed to have a more intense and even more "meaningful" experience than "normal." In other words, there may be some offsetting of the difficulty and unfamiliarity of the task by a heightened capacity and attentiveness. It is clear that more subtle discussion should await results from the follow-up experiments mentioned above, and in particular those in which the listeners do various tests again after having undergone numerous listenings to the whole work and have therefore gotten beyond their initial impressions (hopefully) to more informed and nuanced reactions.[4,5]

References

Aiello, R. (1994). Can listening to music be experimentally studied? In R. Aiello & J. Sloboda (Eds.), *Musical perceptions* (pp. 273–282). New York: Oxford University Press.

Ayari, M., & McAdams, S. (2003). Aural analysis of Arabic improvised instrumental music (taqsîm). *Music Perception, 21,* 159–216.

Balkwill, L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception, 17,* 43–64.

Berlyne, D. E. (1970). Novelty, complexity and hedonic value. *Perception & Psychophysics, 8,* 279–286.

Berz, W. L., & Kelly, A. E. (1998). Perceptions of more complete musical compositions: An exploratory study. *Psychology of Music, 26,* 175–185.

Bigand, E., Vieillard, S., Madurell, F., Marozeau, F., & Daquet, A. (submitted). Multidimensional scaling of emotional responses to music: The effect of musical expertise and excerpts' duration.

Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing.* Chicago: Chicago University Press.

Clarke, E. F., & Krumhansl, C. L. (1990). Perceiving musical time. *Music Perception, 7,* 213–252.

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition, 80,* B1–B10.

Deliège, I. (1989). A perceptual approach to contemporary musical forms. *Contemporary Music Review, 4,* 213–230.

Dember, W. N., & Earl, R. W. (1957). Analysis of exploratory, manipulatory, and curiosity behaviors. *Psychological Review, 64,* 91–96.

DeWitt, L. A., & Crowder, R. G. (1986). Recognition of novel melodies after brief delays. *Music Perception, 3,* 259–274.

---

4. We would like to thank IRCAM-Centre Pompidou for supporting this project: Laurent Bayle, then IRCAM's director, Eric de Visscher, then IRCAM's artistic director, and Hugues Vinet, IRCAM's scientific director made this project possible, and Alain Jacquinot, Emmanuel Fléty, Martine Gaultier and François Gibouin of the Musical Production Department provided invaluable assistance with the conception and construction of the real-time response system. The project also benefited from the collaboration, patience, and good will of Jean-Marie Cottet, Pierre-André Valade, and the Ensemble Court Circuit at the Paris concert, and Gloria Cheng, Harvey Sollberger, and the SONOR Ensemble at the La Jolla concert. James Ramsay of McGill University provided valuable assistance with the functional data analysis. The orientation of participants at the Paris concert was handled by the members of the Music Perception and Cognition team at IRCAM and the Music Cognition team at LEAD. Momilani Ramstrum, Kuei-Jo Lin, Benjamin Carson, and Jason Rosenberg handled the orientation in La Jolla. Ralph Pitt, Ron Quillin, and Josef Kucera graciously helped with the installation at UCSD. The work was funded by the Cognitique program of the French Ministry of Research, and benefited from support from IRCAM-Centre Pompidou and the University of California at San Diego. We would also like to thank two anonymous reviewers and Yoshitaka Nakajima for important critical input to the work.

5. Editor's Note: Guest Editor Daniel Levitin recused himself and Lola Cuddy served as Action Editor for this article because the second author is a student in Levitin's laboratory.

Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology, 1,* 30–49.
Dowling, W. J., Kwak, S., & Andrews, M. W. (1995). The time course of recognition of novel melodies. *Perception & Psychophysics, 57,* 136–149.
Dowling, W. J., Tillmann, B., & Ayers, D. F. (2001). Memory and the experience of hearing music. *Music Perception, 19,* 249–276.
Dubnov, S., McAdams, S., & Reynolds, R. (in press). Structural and affective aspects of music from statistical audio signal analysis. *Journal of the American Society for Information Science and Technology.*
Frederickson, W. E. (1995). A comparison of perceived musical tension and æsthetic response. *Psychology of Music, 23,* 81–87.
Fredrickson, W. E. (1997). Elementary, middle, and high school perceptions of tension in music. *Journal of Research in Music Education, 45,* 626–635.
Gerrig, R. J., & McKoon, G. (1998). The readiness is all: The functionality of memory-based text processing. *Discourse Processes, 26,* 67–86.
Gerrig, R. J., & McKoon, G. (2001). Memory processes and experiential continuity. *Psychological Science, 12,* 81–85.
Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22,* 1166–1183.
Goldinger, S. D., Kleider, H. M., & Shelley, E. (1999). The marriage of perception and memory: Creating two-way illusions with words and voices. *Memory and Cognition, 27,* 328–338.
Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). *Emotional contagion.* New York: Cambridge University Press.
Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention and memory. *Psychological Review, 83,* 323–335.
Jones, M. R., & Yee, W. (1993). Attending to auditory events: The role of temporal organization. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 69–112). Oxford: Oxford University Press.
Juslin, P. N., & Sloboda, J. A. (Eds.). (2001). *Music and emotion: Theory and research.* Oxford: Oxford University Press.
Karno, M., & Konečni, V. J. (1992). The effects of structural interventions in the first movement of Mozart's Symphony in G Minor, K. 550 on aesthetic preference. *Music Perception, 10,* 63–72.
Krumhansl, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas. *Music Perception, 13,* 401–432.
Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology, 51,* 336–352.
Krumhansl, C. L. (1998). Topic in music: An empirical study of memorability, openness, and emotion in Mozart's String Quintet in C Major and Beethoven's String Quartet in A Minor. *Music Perception, 16,* 119–134.
Krumhansl, C. L., & Schenck, D. L. (1997). Can dance reflect the structural and expressive qualities of music? A perceptual experiment on Balanchine's choreography of Mozart's Divertimento No. 15. *Musicae Scientiae, 1,* 63–85.
Lalitte, P., Bigand, E., Poulin-Charronnat, B., McAdams, S., Delbé, C., & D'Adamo, D. (2004). The perceptual structure of thematic materials in *The Angel of Death. Music Perception, 22,* 265–296.
Lerdahl, F. (2001). *Tonal pitch space.* Oxford: Oxford University Press.
Lerdahl, F., & Jackendoff, R. (1983). *The generative theory of tonal music.* Cambridge, MA: MIT Press.
Levinson, J. (1997). *Music in the moment.* Ithaca, NY: Cornell University Press.
Levitin, D. J., Nuzzo, R., & Ramsay, J. O. (submitted). Introduction to functional data analysis.
Madsen, C. K., & Coggiola, J. C. (2001). The effect of manipulating a CRDI dial on the focus of attention of musicians/nonmusicians and perceived aesthetic response. *Bulletin of the Council for Research in Music Education, 149,* 13–22.

Madsen, C. K., & Fredrickson, W. E. (1993). The experience of musical tension: A replication of Nielsen's research using the Continuous Response Digital Interface. *Journal of Music Therapy, 30,* 46–63.

Martindale, C. (1984). The pleasures of thought: A theory of cognitive hedonics. *Journal of Mind and Behaviour, 5,* 49–80.

Matzkin, D. (2001). *Perception de similarité de mélodies tonales et non tonales: Etude pluridisciplinaire [Perception of similarity of tonal and nontonal melodies: An interdisciplinary study].* Unpublished PhD thesis, Ecole des Hautes Etudes en Sciences Sociales, Paris.

McAdams, S., & Battier, M. (Eds.). (2005). *Creation and perception of a contemporary musical work:* The Angel of Death *by Roger Reynolds* [E-book]. Paris: IRCAM-Centre Pompidou.

McAdams, S., & Matzkin, D. (2003). The roots of musical variation in perceptual similarity and invariance. In I. Peretz & R. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 79–94). Oxford: Oxford University Press.

Meyer, L. B. (1956). *Emotion and meaning in music.* Chicago: University of Chicago Press.

Michon, J. A. (1978). The making of the present: A tutorial review. In J. Requin (Ed.), *Attention and performance VII* (pp. 89–111). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nielsen, F. V. (1983). *Oplevelse af Misikalsk Spending [The experience of musical tension].* Copenhagen: Akademisk Forlag.

Palmer, C. (1996). Anatomy of a performance: Sources of musical expression. *Music Perception, 13,* 433–453.

Panksepp, J. (1995). The emotional source of 'chills' induced by music. *Music Perception, 13,* 171–208.

Penel, A. (2000). *Variations temporelles dans l'interprétation musicale: Processus perceptifs et cognitifs [Temporal variations in musical performance: Perceptual and cognitive processes].* Unpublished Unpublished PhD thesis., Université Paris VI, Paris.

Pollard-Gott, L. (1983). Emergence of thematic concepts in repeated listening to music. *Cognitive Psychology, 15,* 66–94.

Poulin-Charronnat, B., Bigand, E., Lalitte, P., Madurell, F., Vieillard, S., & McAdams, S. (2004). Effects of a change in instrumentation on the recognition of musical materials. *Music Perception, 22,* 239–263.

Ramsay, J. O. (2003). *FDA Analysis Toolkit* [unpublished software for MatLab]. Available for download on ftp://ego.psych.mcgill.ca/pub/ramsay.

Ramsay, J. O., & Silverman, B. W. (1997). *Functional data analysis.* New York: Springer Verlag.

Ramsay, J. O., & Silverman, B. W. (2002). *Applied functional data analysis: Methods and case studies.* New York: Springer.

Reynolds, R. (1987). A perspective on form and experience. *Contemporary Music Review, 2,* 277–308.

Reynolds, R. (2001). *The Angel of Death,* for piano solo, chamber orchestra and computer processed sound. New York: C. F. Peters.

Reynolds, R. (2004). Compositional strategies in *The Angel of Death* for piano, chamber orchestra and computer-processed sound. *Music Perception, 22,* 173–205.

Sachs, J. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics, 2,* 437–442.

Schenker, H. (1935/1979). *Free composition* (E. Oster, Trans.). New York: Longman.

Schubert, E. (1996). Enjoyment of negative emotions in music: An associative network explanation. *Psychology of Music, 24,* 18–28.

Schubert, E. (1999). Measuring emotion continuously: Validity and reliability of the two dimensional emotion space. *Australian Journal of Psychology, 51,* 154–165.

Schubert, E. (2000, July). *Unresolved issues in continuous response methodology: The case of time series correlations.* Paper presented at the 7th International Conference on Music Perception and Cognition, Sydney.

Schubert, E. (2001). Continuous measurement of self-report emotional response to music. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 393–414). Oxford: Oxford University Press.

Shepard, R. N., & Teghtsoonian, M. (1961). Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology, 62,* 302–309.

Sloboda, J. A. (1991). Music structure and emotional response: Some empirical findings. *Psychology of Music, 19,* 110–120.

Sloboda, J. A., & Lehmann, A. C. (2001). Tracking performance correlates of changes in perceived intensity of emotion during different interpretations of a Chopin piano prelude. *Music Perception, 19,* 87–120.

Stockhausen, K. (1959–1960). *Kontakte for electronic sounds, piano, and percussion.* London: Universal Edition.

Todd, N. P. (1995). The kinematics of musical expression. *Journal of the Acoustical Society of America, 97,* 1940–1949.

Tulving, E. (1983). *Elements of episodic memory.* Oxford: Oxford University Press.

Vieillard, S., Bigand, E., Madurell, F., & McAdams, S. (2005). Can similarity cues between original and transformed versions of contemporary musical materials produce implicit relations in memory? In S. McAdams & M. Battier (Eds.), *Creation and perception of a contemporary musical work:* The Angel of Death *by Roger Reynolds* [E-book]. Paris: IRCAM-Centre Pompidou.

Vines, B. W., Nuzzo, R. L., & Levitin, D. J. (submitted). Analyzing temporal dynamics in music: Differential calculus, physics, and functional data analysis techniques.

Vines, B. W., Wanderley, M. M., Krumhansl, C. L., Nuzzo, R. L., & Levitin, D. J. (2003). Performance gestures of musicians: What structural and emotional information do they convey? In A. Camurri & G. Volpe (Eds.), *Gesture-based communication in human-computer interaction* (pp. 468–478). Berlin: Springer Verlag.

Ward, L. M. (2002). *Dynamical cognitive science.* Cambridge, MA: MIT Press.

Wörner, K. H. (1963/1973). *Stockhausen: Life and work* (B. Hopkins, Ed. & Trans.). Berkeley: University of California Press.