A Perceptual Analysis Of Mozart's Piano Sonata K. 282: Segmentation, Tension, and Musical Ideas

CAROL L. KRUMHANSL
Cornell University

The experiments reported here provide a perceptual analysis of the first movement of Mozart's Piano Sonata in B♭ Major, K. 282. The listeners, who varied in the extent of their musical training, performed three tasks while listening to the piece as it was reproduced from an expert performance. The first task determined how the music is perceived to be segmented, the second task determined how the experience of tension varies over time, and the third task determined what listeners identify as new musical ideas in the piece. These tasks were performed first on the entire piece and then on smaller sections from the beginning. These three aspects of music perception are coordinated with one another and correlate with various musical attributes. Judgments of section ends co-occurred with peaks in tension and slow tempos. Judgments of new musical ideas co-occurred with low tension levels and neutral tempos. Tension was influenced by melodic contour, note density, dynamics, harmony, tonality, and other factors. Judgments of large-scale section ends were less frequent than judgments of new musical ideas, but these were more nearly one-to-one on smaller time scales. A subsidiary experiment examined the extent to which tension judgments were influenced by performed tempo and dynamics. Listeners made tension judgments for four different versions of the piece: as performed, constant dynamics (with tempo as performed), constant tempo (with dynamics as performed), and constant tempo and dynamics. The tension curves were generally very similar, deviating only in a few regions containing major section ends. The results are considered in light of the metaphor of tension applied to music and the analogy between music and linguistic discourse.

The experience of music is notoriously difficult to describe. As a consequence, a wide variety of different approaches to musical analysis have been developed (as summarized, for example, by Bent, 1987, 1994; Cook, 1987). Each approach has its guiding metaphors, special terminology, descriptive devices, and theoretical commitments. Some are oriented toward specific musical styles or compositional methods and are narrowly focused on musical concerns. Others engage broader philosophical and psychologi-

Requests for reprints may be sent to Carol L. Krumhansl, Department of Psychology, Uris Hall, Cornell University, Ithaca, NY 14853. (e-mail: clk4@cornell.edu)
MUSICAL TENSION

Music, particularly in the Western tonal-harmonic style, is often described in terms of patterns of tension and relaxation (or release from tension). According to Bent (1987), this way of describing music arose in early 20th century musical analysis and was influenced by Gestalt psychology. Three Gestalt principles were extended to music: closure, which automatically completes partially incomplete patterns; the phi phenomenon, which interpolates a link between two separate occurrences; and Pragnanz, which seeks the simplest possible perceptual organization. In addition, figure-ground organization motivated the analysis of music by using reductions to show the essential underlying structure. Musical form was considered a type of whole or Gestalt. Phrases, motives, rhythms, and other musical patterns bear specific relationships with one another, and together determine the musical form. The musical form, in turn, exerts influences on the perception of the musical components, just as a visual form influences the perception of its components. Finally, this tradition of music analysis is congenial with Gestalt theory in its reliance on graphical methods.

One of the leading proponents of this approach was Ernst Kurth. In his terminology, melodic, harmonic, and rhythmic elements carry energy, which drives music forward and defines internally coherent units. Patterns of tension exist at multiple levels, ranging from small-scale features such as chromatic alteration to large-scale features such as shifts in key areas. Rothfarb (1991) cites Kurth’s (1931/1947) *Musikpsychologie* as playing a significant role in defining a music psychology emphasizing learned cultural influences that distinguished it from an acoustically based “tone psychology.” Emphasizing the affinity with Gestalt psychology, he says “...there were so many similarities between Gestalt principles and Kurth’s music-theoretical and cognitive ideas that we could say Kurth intuitively explored in the aural-temporal domain what Gestalists later scientifically explored and experimentally verified in the visual-spatial domain.” Other music analysts have offered similar descriptions of music. According to Meyer (1967, p. 43), “[...m]usic is a dynamic process. Understanding and enjoyment depend upon the perception of and response to attributes such as tension and repose, instability and stability, and ambiguity and clarity” (p. 43). He not only emphasizes the importance of this aspect of the musical experience, but offers some ideas about the psychological basis for it. Expectations are conditioned by processes of pattern perception and learned stylistic habits that lead listeners to expect that certain events will follow. Pattern perception and style knowledge engender strong expectations for continuation at some points in music and at other points they close off units that are complete. Musical meaning and emotion depend on the way in which the actual events in the music play against these expectations, an idea that was developed more fully in Meyer (1956). Related ideas about patterns of tension in music can also be found in the writings of Hindemith (1942), Ratner (1977), Toth (1977), and especially Zuckermandl (1956).

A variety of results in the psychological literature can be seen as supporting the idea that tension is an important dimension of musical experience. For example, Hevner (1936) found that adjectives applied to entire musical selections included a number that are associated with tension: those adjectives contained in her “vigorous-robust” and “exciting-impetuous” clusters of emotion terms. At the other extreme of relaxation were those adjectives contained in the “sentimental-yearning” and “quiet-serene” groups. Similarly, Gabrielson (1973) found one of the major factors distinguishing rhythmic patterns was a dimension with “simple, regular, and clear” at one extreme, and “varied, exciting, and rich in contrasts” at the other extreme. At a more local level, melodic, harmonic, metrical, and rhythmic instability or tension are reflected in a variety of measures, including direct judgments of these qualities, indirect judgments of phrase endings and melodic continuations, and memory and performance errors (see, for example, Bharucha, 1984; Bharucha & Pryor, 1986; Bigand, 1994; Bigand, Parnicutt, & Lerdahl, in press; Krumhansl, 1995; Palmer & Krumhansl, 1987, 1990; and numerous other studies summarized in Krumhansl, 1990,
Graphical expressions of dynamic qualities in musical responses can be found in the work of Trutslit (see Repp, 1993) and Clynies and Nettheim (1982).

The most extensive experimental analysis of musical tension, to my knowledge, however, is the dissertation by Nielsen (1983), which has recently been extended in studies by Madson and Fredrickson (1993; Fredrickson, 1995). According to Nielsen, “In the musical structure of strata, ‘tension’ is assumed to be placed in the middle region of the object, connecting structural characteristics of the surface level with more deeply located strata of emotion and other strata of meaning” (p. 316). In the experiments, listeners pressed a pair of tongs together to indicate the experienced degree of tension. The amount of pressure applied was recorded continuously as the music was played. The musical examples used were the first movement of Haydn’s Symphony, no. 104, and the first 75 measures of R. Strauss’ Also sprach Zarathustra. Listeners were experienced musicians and 16-year-old students. This method produced strikingly regular tension curves, with smaller waves of tension superimposed on larger waves of tension. Intersubject agreement was reasonably strong. Greater consistency was found among the experienced musicians, although some of the students’ tension curves matched those of the musicians. The tension judgments could be related to specific musical factors, including dynamics, timbre, melodic contour, harmony, tonality, and repetition. These factors, Nielsen acknowledges, interact in complex ways and the tension that is felt is assumed to result from the interactions.

Listeners in the experiments to be reported here performed a similar task. As they listened to the movement from the Mozart sonata, they adjusted an indicator on the computer display to show the degree of experienced tension. The position of the indicator was recorded four times per second. The computer control of the music permitted precise registration between the musical events and the responses. This task was repeated a number of times, first with the entire piece, then with smaller subsections. This gives a way of assessing the reliability of the tension judgments across repetitions as the listeners became more familiar with the piece. Listeners had various levels of musical training, so comparisons across individuals examine differences that depend on prior musical experience.

Music and Discourse

Descriptions of music have often relied on analogies between music and language. According to Bent (1994), the earliest surviving full-scale analysis of music, from the Middle Ages, identifies rhetorical sections and other quasi-linguistic units. More recent descriptions have drawn on various parts of linguistic theory, including semiotics (Agawu, 1991; Nattiez, 1990) and formal models of syntax and phonology (Lerdahl & Jackendoff, 1983; Lindblom & Sundberg, 1970; Sundberg & Lindblom, 1976); see Bent (1987, 1994) for additional references and related analytical approaches. The particular parallel between music and language that will be explored here is that between music and discourse. According to this view, music and discourse both consist of units that have well-defined beginnings and ends. Topics are introduced and developed within these units, with various devices used to move the argument forward. Acoustic cues, such as pauses, pitch contour, dynamic stress, and rhythmic patterning, serve to define these units and highlight certain elements within them.

Ratner (1980, 1991) provides a detailed analysis of this parallel between music and discourse in connection with music of the classical style: “music in the early 18th century developed a thesaurus of characteristic figures, which formed a rich legacy for classic composers...They are designated here as topics—subjects for musical discourse” (1980, p. 9). He provides an extensive catalogue of the types of topics or figures found in this style, including dances (e.g., minuet, bourrée, gigue), styles (e.g., military and hunt music, Turkish music, brilliant style, gallant, or free style), and pictorialism or word painting. According to Ratner (1991, p. 616), the opening measures of the E major sonata studied here evoke a wind serene, and the remainder contains references to gigue, German Waltzes, contredanses, sarabandes, polonaises or bourrées, passages in the singing style, the stile legato, and the fantasia style. “Once recognized, they add a final touch of imagery to the coherence and design of tonal patterns. In this process, Mozart, with his incredible skills and his ability to incorporate and synthesize elements from the various styles..., was the greatest master” (Ratner, 1991, p. 619).

On the linguistic side, Chafe (1994) presents a provocative argument for the study of discourse as a way of understanding basic psychological processes. He also summarizes an extensive program of research in discourse analysis and the theory that has developed from it. Very briefly, discourse reflects the internal state of the speaker who brings into consciousness events or objects, often distant in time and place. These are expressed in prosodic units that are the appropriate length to be processed with the aid of echoic memory. Prosody encompasses both physical and perceptual properties, such as pitch, loudness, timing, voice quality, and pauses. Prosodic units begin with what Chafe calls a starting point or point of departure, from which it moves to present new information that is given emphasis through pitch contour, duration, and loudness. Introducing new ideas exacts cognitive costs, so that new ideas are prepared by a context of information that is already active or semiaactive in the consciousness of the speaker (and hearer).
This description shares much with accounts of musical organization. In fact, Chafe says, "Once one has become accustomed to observing intonation units, it becomes impossible not to hear analogous segments in music. Their presence there may be no accident. The convergence of language and music in this respect may very well show a human need to process information in relatively brief units in active consciousness, to combine such units within larger centers of interest, and every so often to shift from one cluster of semi-active information to another" (p. 186). As a demonstration of how his analytic techniques might be applied to music, Chafe presents brief analyses of the beginning of Mozart’s Piano Sonata in F Major, K. 322, and a Seneca Indian song from Drum Dance. The intonational units identified in the analysis appear to be defined primarily by melodic contour and pauses.

Two tasks were included in the experiments reported here to examine aspects of music that may be analogous to discourse. The first aspect is segmentation into hierarchically organized units. One task required listeners to indicate when they heard the end of sections within the piece, similar to the method used by Imberty (1981), Clarke and Krumhansl (1990), and Deliège and El Ahmadi (1990). This was done first for the entire piece, and then for smaller subsections. Of interest are the kinds of features that occur at the ends of sections and whether these changed with the focus on smaller subsections of the music. The second aspect is the introduction and subsequent elaboration of new materials or topics. A second task required listeners to indicate when they heard new musical ideas introduced in the music. Again, this was done first for the entire piece and then for smaller subsections. The responses can be used to determine musical features that set off new musical ideas and can be compared with music analytical concepts such as figure, motif, and phrase (see Bent, 1987). Finally, the positions of the new musical ideas in relation to section endings can be found, as well as how both of these relate to the tension judgments.

Experiments 1–3

The first three experiments all followed the same general method, and the same listeners participated in all three. In the first, listeners heard the entire piece and, on successive hearings, made three different kinds of judgments: section ends, tension, and new musical ideas. Listeners made the same judgments for measures 1–15 in Experiment 2 and measures 1–8 in Experiment 3. The three experiments were always run in the same order, so that experience with the piece increased throughout the series of experiments. The focus in the later experiments on shorter sections might also change the listeners’ thresholds for identifying section ends and new musical ideas and for signaling changes in tension. Comparison across the three experiments allows these three aspects of music perception to be studied at different time scales.

METHOD

Subjects

Fifteen members of the Cornell University community participated as volunteers in the experiment. Their names were entered into a drawing for a gift certificate from Public Radio Music Source. Some of them also received course credit. Listeners varied considerably in their musical ability. At one extreme was a listener who had 3/2 year of formal musical instruction. At the other extreme was a listener with a total of 24 years of instruction summed over a number of different instruments. The median number of years of instruction was 12. Five of the listeners had taken at least one music course at the university level. One listener reported absolute pitch. One reported some previous experience with the piece but had not played it, and one listener had memorized and performed the piece before the experiment.

Apparatus

The music was played under the control of a Macintosh Iicx computer with the MAX software. The input interface connected the computer to a Roland D-20 keyboard synthesizer set to the Acoustical Piano #001 setting (Standard 1-1). The output of the piano was amplified by a Yamaha 1204 MC Series mixing console and presented to listeners over AKG headphones.

Stimulus Materials

The stimulus materials were based on the performance analyzed by Palmer (this issue). The recording on the Bosendorfer 5E acoustic grand piano was coded into MIDI format and then reproduced by the Roland synthesizer. This preserves timing but does not preserve dynamics accurately. During recording, the key velocities rather than actual dynamics were used to code the MIDI velocities, and during reproduction, the dynamics depend on how the synthesizer uses the MIDI velocities. Nonetheless, the reproduction followed the dynamics of the acoustic performance reasonably well, as judged by aural comparison. Experiment 1 used the entire piece of music (which was played with the repeat of the first 15 measures). Experiment 2 used the first 15 measures of the piece, ending with the chord on the first beat of measure 15, which was sounded for two beats. Experiment 3 used the first 8 measures of the piece.

Procedure

The display on the computer screen in Experiment 1 is shown in Figure 1. In Step 1, listeners heard the entire piece so that they could become familiar with it; they made no responses. In Step 2, listeners were instructed to click with the computer mouse on the large button in the center of the screen when they heard the end of each major section of the piece. In Step 3, they adjusted the position of the slider at the center of the screen to indicate the amount of tension; they were asked to try to use the full range of the slider and could adjust it as frequently as they wished. In Step 4, they were instructed to click on the large button in the center of the screen when they heard the start of each new musical idea in the music.

Experiments 2 and 3 were similar, with the following exceptions. Listeners were told that they would perform the same tasks on smaller sections taken from the beginning of the
piece. In Experiment 2, which used the first 15 measures, listeners were asked in Step 2 to indicate the end of each medium-sized section in the piece. After completing Steps 1 through 4, listeners repeated Steps 2, 3, and 4. In Experiment 3, which used the first 8 measures, listeners were asked in Step 2 to indicate the end of each small-sized section of the piece. Again, there was one repetition of Steps 2, 3, and 4.

Before beginning Experiment 1, listeners were given a practice session with a short segment of classical music so that they could become familiar with the interface on the computer and ask any questions they might have about the instructions. This practice session and Experiment 1 took approximately 45 min. Experiments 2 and 3 were run in a separate session lasting approximately 45 min on another day. After completing the three experiments, listeners filled out a questionnaire about their musical backgrounds.

RESULTS

Experiment 1

Large-Scale Segmentation

Listeners were asked to indicate when they heard the end of each major section of the piece. The data were integrated (smoothed) over temporal intervals of two beats because decision and response times produced temporal variability in these judgments. The two-beat interval was selected because integrating over smaller time intervals failed to capture the clustering of responses, making comparisons across tasks and experiments difficult, and integrating over larger intervals unnecessarily reduced temporal precision. Figure 2 shows the percentage of listeners responding within each two-beat interval as a function of time from the beginning of the piece; the gray lines mark the measures. The value shown for the first beat of a measure is the percentage of listeners responding any time between the first and third beats of the measure; the value shown for the second beat of a measure is the percentage of subjects responding any time between the second and fourth beats of the measure, and so on.

The data showed complete consensus among listeners that major sections ended at the end of measures 8, 8 (repeat), 26, and 36 (the end of the piece). Thus, at the highest level, the piece divides into four large segments as indicated by the tree structure shown at the top of Figure 2. There was also fairly strong consensus that a major section ended at the end of measure 21. If the number of listeners responding is taken as an indication of the strength of an ending, then the section extending from measure 9 (repeat) to measure 26 subdivides into the two subsections shown at the next level down in the tree. Measures 15, 15 (repeat), and 33 contained the next largest number of responses, although in the latter two cases, the bimodal pattern found suggests that listeners found two possible ending locations within these measures. The weakest endings were found in measures 3, 13, 3 (repeat), 13 (repeat), and 31, as indicated at the lowest level of the tree.
intervals as the section ends. The duration of each two-beat interval was divided by the duration of the longest two-beat interval in the piece (the interval between the second and fourth beats of the last measure). This is equivalent to dividing the minimum tempo by the local tempo; higher values represent slower local tempos. This ratio is expressed as a percentage in the graph at the bottom of Figure 2. Marked slowing of tempo occurred in measures 8, 8 (repeat), 21, 26, and 36, that is, at the strong section ends. Local variations in tempo also corresponded to the weaker section ends. Two local peaks corresponded to the diffuse section end responses in measures 15, 15 (repeat), and 33, and single peaks corresponded to the section end responses in measures 13, 3 (repeat), 13 (repeat), and 31. Variations of similar magnitude were found in other measures, however, which will be considered again in connection with medium- and small-scale segmentation. Section-end responses correlated significantly with slower tempos \( r = .41 \) \((N = 202, df = 200), p < .0001\).

**Tension**

The second response listeners made was to adjust a slider to indicate the amount of tension heard at each point throughout the piece. Although some listeners adjusted the position of the slider much more frequently than others, most listeners showed the same general patterns. To test this statistically, intersubject correlations were computed. These correlations averaged .42 \((N = 1257, df = 1255), p < .0001\), and ranged from -.03 to .65; all but five of the 105 intersubject correlations were statistically significant (four of which involved the listener with the least musical experience, who moved the slider very frequently). Given this degree of agreement, the remaining discussion will focus on the tension values averaged across listeners.

The average tension values are shown in Figure 3. As can be seen by comparison with the segmentation tree, strong peaks of tension followed by rapid decreases occurred in measures 8, 8 (repeat), and 26, that is, at the strongest section ends. A sharp decrease in tension also occurred at the end of measure 21, which corresponds to the next strongest segment end. Note that the section from measures 16–21 had the most sustained high level of tension. Smaller peaks of tension appeared in measures 3, 13, 15, 3 (repeat), 13 (repeat), 15 (repeat), 31, and 33, which contain the weaker section ends. Thus, tension corresponded quite closely to large-scale segmentation. As would be expected given the correspondence between large-scale segmentation and tempo, tension also corresponded to tempo, shown at the bottom of Figure 3. The largest tension peaks occurred in measures with the slowest tempos, that is, in measures, 8, 8 (repeat), 21, and 26. Smaller tension peaks corresponded to more local variations in tempo in measures 11, 13, 15, 3 (repeat), 11 (repeat), 13 (repeat), 15 (repeat), 18,
Fig. 3. The top graph shows the judgments of tension in Experiment 1. Listeners indicated tension by adjusting an indicator whose position was measured every 250 ms on a scale from 0 to 100. Comparison with the tree structure at the top shows tension peaks followed by rapid decreases at the ends of large-scale sections. Comparison with the bottom graph shows slower tempos at the same points in the music.

Fig. 4. Shows the tension data in more detail. Comparison of the graphs indicates the consistency in the tension responses between repetitions, and between the analogous material in measures 4–15 and 22–33. Theoretical predictions from Eugenio Narmour (personal communication, May 1994) are indicated by $\Theta$ = lower tension and $M$ = higher tension. Structural characteristics producing tension are given in Table 1.

29, 31, and 33. A subsidiary experiment, described later, examines the effect of manipulating tempo on tension judgments.

Figure 4 shows the tension data in more detail. The top and middle graphs show the data for measures 1–15 and 1–15 (repeat), emphasizing the degree of consistency in the responses across repetitions. The material in measures 22–33 is analogous to that in measures 4–15 and, again, the degree of consistency is strong despite the surface differences between these sections. At a general level, the tension judgments correspond to the traditional account presented by Narmour (this issue) of a stable primary theme in measures 1–3 and a stable secondary theme in measures 9–11. However, he notes an alternative interpretation of measures 1–3 as introductory material and measures 4–6 as the primary theme. This is consistent with the relative lack of tension in measures 4–6.

Narmour (personal communication, May 1994) also provided a more detailed account of locations predicted to be low in tension and locations and sources of tension, which are listed in Table 1. As can be seen in the table, the sources of tension are varied, and in many measures, two or more sources work in combination. Figure 4 indicates the locations predicted to be low in tension by $\Theta$, and these correspond to either low tension or are soon followed by decreases in tension. Points of predicted higher tension, indicated by $M$, correspond to either high tension values or are soon followed by increasing tension values. Interestingly, despite the similar tension curves for measures 4–15 and measures 22–33, the sources of tension are not always the same (compare measures 4 and 22, 5 and 23, 6 and 24, and 7 and 25 in Table 1). In sum, the theoretical predictions were confirmed by the listeners' tension judgments.

Comparing the tension judgments with the score suggested that the tension curves might also be influenced by melodic contour. Figure 5 shows the tension curves (leaving out the repetition of measures 1–15) together with a schematic of the notes displayed under the axes. Tension tended to
### TABLE 1
Sources of Tension in Narmour’s Analysis

<table>
<thead>
<tr>
<th>Measure</th>
<th>Source of Tension</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Dissonance</td>
</tr>
<tr>
<td>3</td>
<td>Extraopus harmonic style</td>
</tr>
<tr>
<td>4</td>
<td>Melody, dynamic</td>
</tr>
<tr>
<td>5</td>
<td>Melody, dynamic</td>
</tr>
<tr>
<td>6</td>
<td>Mode</td>
</tr>
<tr>
<td>7</td>
<td>Dissonance, dynamic</td>
</tr>
<tr>
<td>8</td>
<td>Melody</td>
</tr>
<tr>
<td>11</td>
<td>Denial of intraopus style</td>
</tr>
<tr>
<td>13</td>
<td>Denial of intraopus style, denial of extraopus harmonic style</td>
</tr>
<tr>
<td>14</td>
<td>Chromaticism, denial of intraopus style, denial of extraopus harmonic style</td>
</tr>
<tr>
<td>15</td>
<td>Melody</td>
</tr>
<tr>
<td>16</td>
<td>Denial of intraopus style, key change</td>
</tr>
<tr>
<td>17</td>
<td>Chromaticism, dynamic</td>
</tr>
<tr>
<td>18</td>
<td>Denial of intraopus style</td>
</tr>
<tr>
<td>19</td>
<td>Harmonic process, dynamic, chromaticism</td>
</tr>
<tr>
<td>20</td>
<td>Denial of intraopus style</td>
</tr>
<tr>
<td>21</td>
<td>Dissonance, denial of intraopus style</td>
</tr>
<tr>
<td>22</td>
<td>Denial of intraopus style</td>
</tr>
<tr>
<td>23</td>
<td>Break in pattern of harmony</td>
</tr>
<tr>
<td>24</td>
<td>Dissonance, dynamic</td>
</tr>
<tr>
<td>25</td>
<td>Melody, denial of extraopus style, dynamic</td>
</tr>
<tr>
<td>26</td>
<td>Melody</td>
</tr>
<tr>
<td>29</td>
<td>Denial of intraopus style</td>
</tr>
<tr>
<td>31</td>
<td>Denial of intraopus style, denial of extraopus harmonic style</td>
</tr>
<tr>
<td>32</td>
<td>Chromaticism, denial of intraopus style, denial of extraopus harmonic style</td>
</tr>
<tr>
<td>34, 35</td>
<td>Melody, dissonance</td>
</tr>
</tbody>
</table>

Fig. 5. Tension curves together with a schematic diagram of the notes displayed under the axes. Some of the fine-grained detail of the melodic contour is reflected in the fine-grained detail of the tension judgments. Decreased density of notes at the ends of major sections correspond to the largest drops in tension. The figure also shows a schematic diagram of the MIDI velocities that correspond approximately to dynamics. Some of the fine-grained detail of the dynamics is also reflected in the fine-grained detail of the tension judgments. Softer dynamics at the ends of major sections correspond to decreases in tension.

#### Musical Ideas

The last task listeners performed was to indicate the start of each new musical idea in the piece. As with the judgments of section ends, these judgments exhibited a degree of temporal variability that motivated integrating the responses over two-beat intervals. Figure 6 shows the percent-

---

covary with the pitch height of the melody at a local level (the highest notes in the schematic). This was particularly so in measures 9–15, 16–19, and 27–33, where the fine-grained detail of the tension curves quite closely followed the melodic contour. This figure also makes clear that the sharpest drops in tension occurred when the density of notes decreases in measures 8, 21, 26, and 36. Figure 5 also displays the MIDI-coded key velocities under the tension graphs. As noted earlier, these values only approximate the dynamics of either the original performance or the version that was reproduced in the experiment. Even so, some correspondence was found. The fine-grained detail in tension in measures 9–15 and 27–33 tended to follow the highest velocity values. Also, the major section ends in measures 8, 15, 21, 26, and 36 were accompanied by declines in dynamics. Thus, influences of both dynamics and pitch height can be found in the tension profiles. It is interesting to note the correspondence between dynamics and pitch height suggested by this figure; the correlation between MIDI velocities and pitch height was $r = .51$ ($N = 1540, df = 1538$), $p < .0001$. This is noteworthy given the dissociation found by Palmer (this issue) between these variables as they correlate with theoretical predictions. The subsidiary experiment reported later investigates the effect of manipulating dynamics on perceived tension.
correlation of \(-0.05\) \((N = 202, df = 200)\), which is not significant. An interesting relationship held between the new musical ideas and tension, which can be seen by comparing Figures 4 and 6. New musical ideas tended to occur at positions where tension was relatively low or had just declined markedly. This is the pattern found in measures 1, 4, 9, 11, 14, 15, 1 (repeat), 4 (repeat), 9 (repeat), 11 (repeat), 14 (repeat), 15 (repeat), 16, 22, 27, 29, 32, 33, and 34. The only exceptions are the responses of new ideas that occurred in measures 19 and 20, although even these co-occur with a local decrease in tension.

Figure 6 also shows an analysis by Robert Gjerdingen (personal communication, March 1995) of how the piece might be segmented according to musical figures. The beginning of each figure is indicated in the graph by an asterisk. With the one exception of measure 6 (repeat), these locations were followed by judgments that new musical ideas had occurred. Thus, good agreement was found between the predictions and listeners' judgments of new musical ideas, suggesting that the judgments were based on a figurative strategy of segmentation. Gjerdingen (this issue) describes the historical precedents that listeners might have associated with some of these figures: Opening Gambit in measure 1 (his Figure 4), Prinzer Riposte in the middle of measure 2 (his Figure 4), and Fonte in measure 16 (his Figure 4).

Comparison with the score shows that the new-idea judgments corresponded to multiple surface characteristics. The melody exhibited a change in rhythm and a new pitch pattern, which in most instances was repeated almost immediately. Often, a predominantly descending melodic pattern changed to an ascending pattern or, less frequently, the opposite. Sometimes what listeners took to be a new musical idea was also signaled by a shift in the register of the melody. The accompaniment also marked new ideas in various ways: by introducing new rhythmic and pitch patterns, changing the density of the material, and shifting register. Less often, new musical materials corresponded to changes in dynamics. In all cases, multiple surface characteristics worked in combination.

Finally, whenever previously heard material was reintroduced, judgments of new musical ideas occurred. Consistent with this, listeners occasionally commented that they interpreted the instructions to mean that they were to indicate a change in the musical idea even if it had been presented earlier in the piece.

Experiments 2 and 3

Medium- and Small-Scale Segmentation

Experiment 2 used the first 15 measures of the music, and listeners were asked to judge where endings of medium-sized sections occurred. As before, the responses were integrated over two-beat intervals. The percent-
which had previously elicited few responses, now received more responses, and a few additional responses occurred in measure 5. Again, even with these additional responses, significant correlations were found with the other experiments: \( r = .71 \) (\( N = 32, df = 30 \)) with Experiment 1, and \( r = .92 \) (\( N = 32, df = 30 \)) with Experiment 2, both significant at \( p < .0001 \).

Although increased familiarity with the piece and the focus on the smaller sections of the piece generated more responses, the timing of the responses that occurred in all three experiments was approximately the same. That is, listeners did not respond more rapidly in the later experiments. Also, in both Experiments 2 and 3, the judged locations of endings corresponded to tempo. The additional responses on the smaller scale picked up tempo variations that had not been accounted for in Experiment 1. The correlation with tempo was \( r = .51 \) (\( N = 57, df = 55 \)) in Experiment 2, \( p < .0001 \), and \( r = .39 \) (\( N = 32, df = 30 \)) in Experiment 3, \( p = .03 \).

**Tension**

Figure 8 shows the tension ratings for Experiments 1, 2, and 3. Although the later experiments generated profiles with slightly more fine-grained structure, the patterns were remarkably similar. The correlations between the experiments using the tension data averaged across listeners were: Experiments 1 and 2, \( r = .97 \) (\( N = 333, df = 331 \)); Experiments 1 and 3, \( r = .96 \) (\( N = 185, df = 183 \)); and Experiments 2 and 3, \( r = .99 \) (\( N = 185, df = 183 \)), all significant at \( p < .0001 \).

In addition to the strong consistency across experiments, strong consistency was found within experiments across repetitions. In both Experiments 2 and 3, listeners repeated the tension task twice. The average correlation between repetitions for individual subjects was \( r = .78 \) (range, .53 to .91; \( N = 333, df = 331 \)) in Experiment 2, and \( r = .81 \) (range, .53 to .98; \( N = 185, df = 183 \)) in Experiment 3. All 30 replication correlations were significant at \( p < .0001 \). In addition, as in Experiment 1, intersubject correlations were also strong. In Experiment 2, they averaged .47 (\( N = 333, df = 331 \)) and ranged from .06 to .88; all but six of the 105 correlations were statistically significant at \( p < .05 \). In Experiment 3, they averaged .46 (\( N = 185, df = 183 \)) and ranged from .30 to .92; all but six of the 105 correlations were statistically significant. Nine of the 12 nonsignificant correlations involved a single subject who adjusted the slider much more frequently than she had in Experiment 1, perhaps responding to the instructions to focus on smaller sections.

Because of the strong agreement between the tension data in Experiments 1, 2, and 3, the same correspondences to tempo, melodic contour, and dynamics would be expected. In fact, the more fine-grained structure in the later experiments tracked these features of the music even more closely.
than before, as can be seen by comparison with Figures 3 and 5. In addition, tension on the medium- and small scales was even more tightly coupled with judgments of section ends than in Experiment 1. Peaks in tension coincided with section-end judgments in measures 2, 3, 6, 7, 8, 13, and 15 in Experiment 2, and additionally in measure 5 in Experiment 3. The only disparity was in measure 11.

The tension ratings for the first eight measures were averaged over the three experiments and are shown in Figure 9. For this opening segment, Lerdahl (this issue, Figure 21) provided a set of numerical predictions for tonal tension of each event. The definitions of the terms and the rationale for their numerical coding appear in his article. Briefly, scale degree codes whether the melodic tone is contained in the supporting triad. Inversion codes whether the triad is in root position or inversion. Nonharmonic tone codes for the presence of tones not in the chord, including chordal sev-

ents. These three variables are considered sources of surface dissonance. The remaining variables come from Lerdahl's (1988) pitch-space model. A chord is notated by its root position in a designated reference key, for example, E = I of B minor. Pitch-space i distance between two chords, x and y, is the distance between them along the cycle of fifths operating at the diatonic level (i.e., the number of diatonic scale tones changed between the reference keys). Pitch-space k distance is the distance between the chords along the cycle of fifths for chords (I – V – ii – vi – iii – vii – IV – I). Pitch-space k distance is the number of tones that are not shared by the two chords counted at all levels of the basic pitch space (Lerdahl, this issue, Figure 4). In essence, this method of counting weights the k distance according to the scale position of the unshared tones: greatest weight for the tonic, next for the dominant, next for the third, and lowest for other diatonic scale tones. Finally, inhe...
value is the sum of the pitch space \( i, j, \) and \( k \) values for all events superordinate to an event in the prolongational tree. This gives a total of seven quantitative variables.

The fit of the tension judgments by the model with these seven variables was assessed by using multiple regression. The multiple regression was highly significant \( R^2(7,177) = .79, p < .0001 \), indicating a good fit to the data. Subsequent regression analyses showed that the surface dissonance variables accounted for the data less well \( R^2(3, 181) = .28, p = .001 \) than the four pitch-space variables \( R^2(4, 180) = .78, p < .0001 \). Indeed, these four variables alone accounted for the data as well as all seven variables together. Of the pitch-space variables, the strongest was inherited value (see also Palmer, this issue), followed by pitch-space \( k \) distance.

The top of Figure 9 compares the tension judgments with the full seven-variable model. As can be seen, the responses tended to lag slightly behind the theoretical predictions. That is, a rise in predicted tension is often followed shortly by a rise in perceived tension and similarly for drops in tension. In addition, the model predicts more fine-grained detail than is found in the judgments. Together, these results suggest that listeners are integrating the musical information over time, producing the smoother and slightly delayed tension profiles. To test this, a model that included lags in units of 250 ms from 0 to 3250 ms (approximately two beats of the music) was tested. The fit of the model is shown at the bottom of Figure 9. This model with lags provided a considerably better fit to the data \( r(14,170) = .91, p < .0001 \). The smoothness of the tension judgments suggests that they would not be modeled well by either Narmour’s (this issue, Figures 25–27) values of closure and nonclosure or the strength of Bharucha’s (this issue) yearning vector, which appear to apply on a more local time scale.

Musical Ideas

Finally, Figure 10 shows the percentage of listeners identifying new musical ideas in each two-beat interval during the initial segment of the piece in Experiments 1, 2, and 3. Comparing first the results for Experiments 1 and 2, we see similar patterns. However, the responses tended to occur somewhat earlier in Experiment 2 than in Experiment 1. In part because of this, the correlation between Experiments 1 and 2 was negative \( r = -.23 \) \((N = 57, df = 55)\), which approached significance, \( p = .08 \). However, shifting the Experiment 1 data earlier by three beats produced a good match between the experiments \( r = .75 \) \((N = 57, df = 55), p < .0001 \). This temporal lag corresponds approximately to the time it takes for the new material in the melody and/or accompaniment to be repeated, which may be necessary in the first experiment to reinforce the impression that a new musical idea has been introduced. In addition to the temporal shift in responding in Experiment 2, more subjects responded to the new material in measures 6, 7, and 12 than before.

Experiment 3 was similar to Experiment 2, even though more listeners gave responses in measures 6 and 7, and some new responses appeared in measure 2. These experiments correlated quite strongly with one another \( r = .71 \) \((N = 32, df = 30), p < .0001 \). However, the temporal shift in the responses relative to Experiment 1 was again apparent and produced a negative correlation \( r = -.33 \) \((N = 32, df = 30)\) between Experiments 1 and 3, which again approached significance, \( p = .07 \). In this case, shifting by five beats produced the highest correlation between these experiments \( r = .35 \) \((N = 32, df = 30), p = .05 \).

Fig. 10. Judgments of new musical ideas in Experiments 1, 2, and 3. Focusing on smaller sections produced responses in more locations. In addition, responses occurred more rapidly in Experiments 2 and 3. The difference was approximately one measure relative to Experiment 1. Comparison with Figure 7 shows a close correspondence between judgments of new musical ideas and section ends on the smaller scales. Comparison with Figure 8 shows that new musical ideas corresponded to relative low levels of tension.
Comparison with Figure 7 shows an increasingly close correspondence between the locations of judged endings and the introduction of new musical ideas. In the later experiments, judgments of section ends were always followed shortly by judgments of new ideas. These two variables had a correlation of \( r = .59 \) (\( N = 57, df = 55 \)), \( p < .0001 \) in Experiment 2, and \( r = .45 \) (\( N = 32, df = 30 \)), \( p = .01 \) in Experiment 3. This suggests that the smaller scale sections are defined largely by a figural strategy of segmentation. Comparison with Figure 8 shows the relationship between the locations of new musical ideas and tension was also stronger in these experiments. New musical ideas tended to be identified at points in the music where the tension level was either low or had just declined markedly. Indeed, the additional judgments of musical ideas that appeared in these later experiments in measures 2, 6, and 7 can be linked to drops in the tension values in these regions. Only those in measures 11 and 12 did not follow this pattern. Finally, new musical ideas tended to be introduced when tempo was at a neutral level. The correlations of these variables were \( r = .24 \) (\( N = 57, df = 55 \)) and \( r = .05 \) (\( N = 32, df = 30 \)) in Experiments 2 and 3, respectively, neither of which was statistically significant. In sum, on the smaller time scales, judgments of new musical ideas quite consistently followed section ending judgments and occurred at points of low tension and neutral tempo.

**Experiment 4**

Before turning to a discussion of the results, a fourth, subsidiary experiment will be presented. The first three experiments revealed a number of relationships between segmentation, tension, and musical ideas. In addition, some of these correlated with performed variations in tempo and dynamics. This raises the question as to the causal nature of the links between performance variations and perceptual responses. How would the responses change for a temporally regular or dynamically level performance? Listeners in the fourth experiment heard four different versions of the piece: as performed, constant dynamics (with performed tempo), constant tempo (with performed dynamics), and constant dynamics and tempo. In the interest of time, listeners made tension judgments only. This task was selected because it seemed intuitively to be the most susceptible to performance nuances. In contrast, segmentation and musical ideas would seem to be signaled by numerous cues contained in the notated pitches and durations independently of how the piece is performed.

**METHOD**

**Subjects**

Twenty-four members of the Cornell community participated in the experiment for which they received course credit. Listeners had from 1 to 18 years of instruction on musical instruments, with a median of 10 years. Five had taken at least one music course at the university level. Two reported absolute pitch. One reported knowing the piece from a recording they owned, and three said they thought they recognized it but were not sure.

**Apparatus**

Same as in Experiments 1, 2, and 3.

**Stimulus Materials**

All versions of the piece used in this experiment contained only one presentation of measures 1-15. With this difference, the first version was the same as Experiment 1. (To make a natural sounding beginning, the durations and velocities from the beginning of the original performance were used at the beginning of measure 1; otherwise, the values were based on the performance of the repetition.) The second version, constant dynamics (performed tempo), used the same timing as the first but a constant MIDI velocity (100), corresponding to a moderately loud dynamic. The third version, constant tempo (performed dynamics), used the original MIDI velocities, but a constant tempo. The duration of the entire piece was the same, and all tone onsets were adjusted to correspond to the notated durations. The fourth version, constant dynamics and tempo, used the constant dynamics of the second version and the constant tempo of the third.

**Procedure**

The display on the computer screen contained only the slider and instructions for making the tension judgments. Listeners started with a similarly adapted practice session. Each listener heard all four versions, the order of which was determined by a Latin square so that each version was heard equally often in each position. Afterward, the listeners filled out the questionnaire. The experimental sessions lasted approximately 45 min.

**RESULTS**

The main focus will be on the tension judgments averaged across listeners. Preliminary analyses showed that intersubject correlations in this experiment were lower than previously, averaging \( r = .18 \) (\( N = 897, df = 895 \)), which, however, is still significant at \( p < .0001 \). The average intersubject correlations were approximately equal for the four versions: as performed, \( r = .18 \); constant dynamics (performed tempo), \( r = .19 \); constant tempo (performed dynamics), \( r = .19 \); and constant dynamics and tempo, \( r = .16 \). Nor were there obvious differences depending on the order in which the versions were presented. Consequently, the data were averaged across listeners (and, consequently, the order of presentation). Figure 11 shows the tension judgments for the four versions, which contain remarkably similar patterns. Correlations can only be computed between versions with the same timing patterns. The correlation between the performed version and the constant dynamics (performed tempo) version was \( r = .88 \) (\( N = 897, df = 895 \)), \( p < .0001 \). These values are shown in the first and third graphs. The only notable difference is the slightly more rapid drop in tension at the major section ends in measures 8 and 26. The correlation between the constant tempo (performed dynamics) version and the
One of the metaphors explored in this study was tension as an aspect of musical experience, which was measured by tracking responses continuously during the reproduction of an expert performance. The other metaphor was music as discourse, which was assessed by soliciting judgments of segmentation and new musical ideas. The primary methodological question explored here was whether listeners can make reliable and interpretable responses without interrupting or otherwise artificially manipulating the musical materials. In other words, is it possible to probe the perception of music as it would normally be experienced? All three kinds of responses elicited from listeners showed precise time-locking to musical events and considerable reliability across repetitions. Only the third task, identifying new musical ideas, exhibited a temporal change in responding with increased experience with the piece; new ideas were identified somewhat more slowly the first time the task was performed.

Another methodological question explored in the study was the utility of using a number of different methods with the same piece of music. How would the different aspects of perception correspond to one another? In addition, would the correspondences change as listeners heard progressively smaller sections of the piece? Figure 12 shows the correspondences that held for the smaller time scales. The different tasks had rather consistent relationships with one another and with tempo. Judgments that a sec-

![Graph showing tension judgments from Experiment 4](image)

Fig. 11. Tension judgments from Experiment 4, which manipulated dynamics and tempo. Listeners heard four versions of the music: as performed, constant dynamics (with performed tempo), constant tempo (with performed dynamics), and constant dynamics and tempo. Only one repetition of measures 1–15 was played. The four tension curves showed remarkably similar patterns, with deviations apparent only at the ends of some of the major sections.

constant dynamics and tempo versions was $r = .83$ ($N = 897, df = 895, p < .0001$). These values are shown in the second and fourth graphs. These two differ primarily in measures 21 and 26, where the dynamics seem to enhance the large drop in tension. Comparison between the versions with different tempos is more difficult. However, visual inspection shows the graphs have very similar shapes, with nearly equal average values, ranges, and degrees of variation for all four versions. In general, it would seem that the manipulation of performed tempo and dynamics had remarkably little effect on the tension judgments.

**Discussion**

The discussion will focus on a number of issues of experimental methodology, and the primary empirical results will be reviewed in that context.
tion end had occurred were soon followed by judgments that a new musical idea had been introduced (and, conversely, judgments that a new musical idea had occurred were almost always preceded by judgments that a section end had occurred). Section-end responses corresponded to peaks in tension followed by rapid decreases in tension and slower tempos. In contrast, new musical ideas were introduced when tension was at a low level and the tempo was neutral. On the large scale, the relationships between these variables were not quite as one-to-one. In particular, listeners judged there to be more new musical ideas than major section ends. Consequently, judgments of section ends were always followed by judgments of new musical ideas but judgments of new musical ideas also occurred within large-scale segments. Employing the three tasks in combination showed these three aspects of music perception are generally quite tightly coordinated with one another.

Another methodological issue considered here was the influence of individual differences in musical experience. Does this cause listeners to respond differently from one another? Listeners in these experiments varied considerably in musical training, although none would qualify either as a professional musician or as a total novice. It is difficult to assess musical expertise and aptitude, but indicators such as years of formal instruction and extent of academic training showed considerable variability. Indeed, one of the listeners had, previously to the experiment, memorized and performed this particular sonata. Despite these differences, responses in the experiment were quite uniform. The nature of two of the tasks, the judgments of section ends and new musical ideas, made it difficult to test statistically the agreement between listeners. Nonetheless, considerable consensus was apparent, particularly about the locations of section ends. Evidently, these aspects of musical structure are expressed quite explicitly in the perceptual information. It was possible to examine statistically the degree to which listeners agreed with one another on their judgments of musical tension. Strong intersubject agreement was found, with no consistent relationship with musical training or other aspects of their musical backgrounds. Also, the tension judgments were highly reliable over repetitions, and changed little with increased experience with the piece over the course of the experiments.

Finally, the first three experiments raised the question of how strongly the performed dynamics and tempo influenced the perceptual judgments. These experiments showed that lower tension ratings tended to co-occur with lower dynamics and, as noted earlier, higher tension ratings tended to co-occur with the slower tempos. To examine this question, versions of the piece were created with constant dynamics and tension and were presented to listeners with the tension task. These manipulations produced remark-ably little change in the tension judgments. To the minimal extent that differences were found, they were noticeable at points with large variations in tension. It would seem, then, that these aspects of performance are not necessary for the listener to experience variations in tension. Instead, it appears that tension is conveyed by the pitch and durational patterns in the music, to which both listeners and performers respond.

Tension is one of many metaphors that has been offered for the musical experience. The results of these experiments, and those of Nielsen (1983) and Madson and Fredrickson (1993; Fredrickson, 1995), suggest that it is amenable to experimental measurement. The present findings exhibited two general patterns of possible interest. First, peaks in the tension ratings tended to be asymmetric. In most cases, tension increased gradually and decreased rapidly. This may be analogous to the tendency for melodic contour to increase in small steps and to decrease in larger steps. Consistent with this, the experiments found that tension judgments covaried somewhat with melodic contour. Second, local peaks of tension were superimposed on larger variations. Local variations of tension correlated with both melodic contour and dynamics, whereas larger scale variations were associated with harmonic instability and shifting tonality.

These results concerning musical tension raise a number of questions. For example, are some features of the tension data in these experiments characteristic of the musical style, and might they be found for music in other styles also? Does tension relate to some aspect of the difficulty of production, such as the physical demands of singing in higher registers and at louder dynamics? Analogies of this sort have been suggested, for example, by Sundberg (1987). Does perceived musical tension relate to patterns of movement by performers, conductors, or dancers? Finally, and most obviously, is there some connection between perceived tension and emotional responses to music? Do different patterns correspond to different emotions? For instance, would psychophysiological measures exhibit similar variations?

The second metaphor explored in this study was the analogy between music and discourse. Music in the classical style has often been compared with discourse. Both are segmented into units within which ideas (or topics) are introduced and developed. Various findings in these experiments encourage further exploration of this analogy. Even though the listeners in this study were doubtless largely unaware of the kinds of musical references described by Ratner (1980, 1991) and Gjerdingen (this issue), there was considerable agreement in identifying new musical ideas. Comparison with the music showed that new musical ideas were marked by a variety of surface characteristics, such as changes in rhythmic and pitch patterns, register, and texture. Moreover, new musical ideas tended to be introduced
after segment boundaries. This was particularly true for segmentation as judged on the smaller scale, where segmentation corresponded mostly closely to the introduction of new musical ideas.

The framework for discourse analysis proposed by Chafe (1994) brings out a number of other similarities. First, new musical ideas tended to be introduced at points of low tension and neutral tempo, which may correspond to his starting points or points of departure that are prepared by the larger context. Second, section ends identified by listeners at all levels in these experiments corresponded to slowing of tempo, perhaps analogous to the patterns of phrase final lengthening and pauses at the end of discourse units. Third, section ends, like the ends of prosodic units, tended to be marked by descending contour and decreased dynamics. Finally, the asymmetric patterns of tension within segments, noted above, may correspond in some way to how ideas are developed and completed within linguistic units.

Again, many unanswered questions remain. Do repetitions in music relate in some way to the units of semiaactive information described by Chafe? Is there a quantifiable correspondence between the cognitive demands of musical and linguistic units? Their durations? Are comparable patterns found in other pieces, or other musical styles? Whatever the answers to these questions may be, the results of the experiments presented here suggest that this particular piece of music coordinates a number of different perceptual and conceptual structures in a way that invites comparison with linguistic discourse.

Comparisons between the experimental results and the theoretical analyses of the piece (Gjerdingen, this issue; Lerdahl, this issue; Narmour, this issue) also raise a number of questions. However, the numerous points of convergence suggest an increasing understanding of the musical structures that underlie music perception. One analysis by Gjerdingen that described how the piece divides into distinctive figures corresponds well to the new musical ideas identified by listeners. Although contemporary listeners are unlikely to have the associations to historical precedents described by Gjerdingen (this issue), they nonetheless appear able to identify the appropriate figural constituents. Narmour’s (this issue) description of the formal design of the piece corresponded to listeners’ segmentation judgments, and his qualitative analysis of sources of tension in the music corresponded to listeners’ tension judgments. Finally, Lerdahl’s (this issue) quantitative predictions of tonal tension provided a good model of the tension judgments in the opening section of the piece. The success of the model supported both local effects of harmonic tension and more global influences depending on an event’s position in the proposed hierarchical tree.

Convergence of this sort with the perceptual data provides external validation for the experimental methods. In turn, the perceptual data help clarify some of the theoretical observations. Such comparisons are also useful for refining questions about psychological processes and suggesting musical structures that might be manipulated experimentally. For example, insofar as dissonance influences perceived tension, is it dissonance of a sensory nature or dissonance established within the context of a particular musical style? How much knowledge of the style is necessary to apprehend the form of a piece and to identify and remember its major constituents? How would composers’ styles or historical changes in performance practice alter the perceptual representation of the music? The present focus on a single piece, indeed a single performance of that piece, imposes severe limits to the generality of the results. One advantage of the approach, however, is that it can suggest ways in which the different approaches inform one another.1

References


1. This research was begun while the author was a Fellow at the Center for Advanced Study in the Behavioral Sciences, Stanford, CA 94305, with support from the National Science Foundation (SES-9022192) and the James McKean Cattell Fund. I would like to express my gratitude to Fred V. Nielsen for a copy of the dissertation, Oplevelse af musikalsk spanning (The experience of musical tension) and to Alf Gabrielsson for bringing it to my attention. I also extend my thanks to Peter Tender at Ohio State University for providing the MIDI code for the music used in the experiment, to Robert Gjerdingen, Fred Lerdahl, and Eugene Narmour for their theoretical analyses of the piece, and to Fred Horan and Carole Lunney for technical assistance. Special acknowledgment goes to Jaimeh Jemison, who conducted the experimental sessions for the first three experiments, carried out many of the preliminary statistical tests, and offered many useful observations. Christopher Lang edited the MIDI code for the last experiment, conducted the experimental sessions, and performed preliminary statistical tests. I am grateful to Bruno Repp for numerous helpful suggestions on an earlier version of this manuscript.
Anatomy of a Performance: Sources of Musical Expression

CAROLINE PALMER
Ohio State University

Research on music performance often assumes that a performer's intention to emphasize musical structure as specified in a score accounts for most musical expression. Relatively unstudied sources of expression in performance include notational variants of compositional scores, performer-specific aspects, and the cultural norms of a particular idiom, including both stylistic patterns that exist across musical works and expectations that arise from a particular musical context. A case study of an expert performance of a Mozart piano sonata is presented in which influences of historical interpretations of scores, performer-specific treatments of ornamentation and pedaling, and music-theoretic notions of melodic expectancy and tension-relaxation are revealed. Patterns of organization internal to the performance expression transcended the coarse-grained information given in scores, suggesting that performance is a better starting point than a musical score for testing theories of many musical behaviors.

DESPITE the fact that music performance provides a rich perspective on our musical experiences, our understanding of music performance has not caught up with empirical study of other types of musical experiences. Take for example the empirical research published in this journal (one of the leading journals in the study of music cognition) in the past 5 years: fewer than one fifth of the articles address performance, and twice as many address perception of music. Some of this disparity arises from the fact that only a minority of people perform music formally in our culture, whereas almost everyone listens to music. Another source of the disparity arises from the imperfect (or absent) methodologies for studying performance; a single performance typically results in a large quantity of complex measurements, and analysis techniques have been lacking. Recently, techniques for measuring and quantifying performance have improved with the advance of computer-aided musical instruments.

Requests for reprints may be sent to Caroline Palmer, Psychology Dept., 1885 Neil Ave., Ohio State University, Columbus OH 43210. (e-mail: cpalmer@magnus.acs.ohio-state.edu)