

Memory for musical surface

CAROL L. KRUMHANSL
Cornell University, Ithaca, New York

This study of memory for music presented listeners with the first half of a short piano piece, O. Messiaen's *Mode de valeurs et d'intensités*. The piece is written according to a unique compositional principle that rigidly couples values of pitch (chroma and octave), duration, and dynamics. Listeners heard test excerpts, which were judged in terms of whether or not they might have come from the piece (either from the part they had heard or from the remainder of the piece). Even in the first block of trials, listeners were able to recognize segments from the part of the piece they had heard, suggesting surprisingly accurate memory for surface characteristics. Listeners were also able to generalize to the rest of the piece, accurately judging segments from the part of the piece they had not heard. However, performance on four kinds of transformed segments showed that the abstraction of a piece's surface characteristics adopts rather loose criteria. Even highly trained professional musicians informed of the compositional constraints before the experiment were not sensitive to the specific couplings between musical parameters. Rather, listeners appeared to be sensitive to contour (the pattern of increasing and decreasing pitch) and global correlations between pitch height and duration, dynamics, and interval size. Performance did not change with repeated hearings of the music.

This study is an exploratory investigation of memory for surface characteristics of music. Specifically, it addresses three interrelated issues: (1) the capacity to recognize music written in an unfamiliar style, (2) the ability to abstract the piece's surface characteristics and generalize to new musical passages, and (3) the musical features important for the process of abstraction and generalization. Psychological studies of music (summarized, for example, in Dowling & Harwood, 1986; Handel, 1989; Krumhansl, 1990, 1991) have concentrated on aspects of musical organization typical of broad classes of musical pieces, rather than characteristic aspects of particular composers or pieces. Western music from the 18th century to the present, for example, tends to use major and minor diatonic scales and triadic harmonies with conventional chord progressions and cadences. Knowledge of these interopus regularities has been referred to as *musical schemata* (Dowling, 1978; Krumhansl & Castellano, 1983; Shepard & Jordan, 1984). In numerous experiments, these schemata have been shown to influence how musical sequences are encoded and remembered.

In one study, for instance, Dowling (1978) found listeners could not reliably distinguish between two transformations of a melody. One transformation was an exact transposition (changing the key, but maintaining interval sizes); the other transformation was a transposition up or down the scale of the original melody (maintaining the

key, but changing the interval sizes). The result suggests that the notes of the melody are assimilated to schematic knowledge of scale structure. Krumhansl (1979) found that nondiatonic tones (those not contained in the scale of the key) are difficult to remember and relatively frequently confused with diatonic tones. In contrast, Cuddy, Cohen, and Miller (1979) and Krumhansl (1979) found that changes from diatonic to nondiatonic tones were relatively easy to detect. Asymmetries of this kind have also been reported by Dowling and Bartlett (1981; Bartlett & Dowling, 1988). Similar results showing assimilation to compositional norms have been found for chord sequences (Bharucha & Krumhansl, 1983) and rhythmic patterns (Bharucha & Pryor, 1986; Palmer & Krumhansl, 1990).

This research has been useful for demonstrating that musical sounds are interpreted in terms of knowledge of compositional regularities underlying a broad class of musical pieces. The rationale for the research has been to obtain results that have general applicability. However, this focus may have underestimated our sensitivity to surface characteristics of music. These surface characteristics may be unique to a particular piece of music, or even a particular performance of that piece. A piece may be distinctive in its employment of melodic gestures, rhythmic figures, orchestration, or phrasing. It seems that listeners may intuitively have the ability to encode and remember large amounts of musical detail given the number of pieces that can be recognized, the speed with which a singer can be identified, and the precision with which timing, dynamics, and other performance gestures can be anticipated. Moreover, listeners learn to recognize a composer or performer in works they may be hearing for the first time. At present, little is known about the psychological processes underlying these abilities.

This research was supported by Grant INT-8701948 from the National Science Foundation. I would like to thank Eric F. Clarke for technical assistance and advice on experimental design, and Howard Kurtzman for useful discussions. Reprint requests should be addressed to Carol L. Krumhansl, Department of Psychology, Uris Hall, Cornell University, Ithaca, NY 14853.

A similar distinction between surface and interpreted memory representations appears in research on language (Anderson, 1974; Anderson & Paulson, 1977; Garrod & Trabasso, 1973). Numerous studies (e.g., Bransford, Barclay, & Franks, 1972; Jarvella, 1971; Sachs, 1967; Wanner, 1968) have shown that subjects remember the meaning or gist of a sentence and not its surface structure—that is, the exact way the meaning is expressed. It is assumed that verbatim memory is short-term and that long-term memory consists primarily of abstract, propositional codes for meaning. However, under some circumstances, verbatim memory can be quite accurate. James, Thompson, and Baldwin (1973) and Graesser and Mandler (1975) showed significant memory for surface structure when subjects were instructed to attend to wording. Even without such explicit instructions, memory for certain kinds of surface detail was quite accurate. Kintsch and Bates (1977) and Keenan, MacWhinney, and Mayhew (1977) showed good surface memory for items in natural discourse, such as jokes, asides, and figures of speech. Bates, Masling, and Kintsch's (1978) recognition data showed effects of the relative explicitness or discourse markedness of a given sentence form; Offir's (1973) data showed effects of presuppositional information. As summarized by Bates, Kintsch, Fletcher, and Giuliani (1980), these results suggest that the strength of verbatim memory depends on the role of the sentence in the discourse. Certain kinds of sentences carry important information in the particular way the meaning is expressed, and surface memory for these tends to be relatively accurate.

This research on memory for language suggests that surface memory for music may also be quite accurate. In addition to the more abstract representations that relate the sounded materials to schematic musical structures, listeners may remember surface characteristics. For the present study, which explores the possibility, it seemed important to use music written in an unfamiliar style, so that the memory judgments would necessarily be based on memory for surface characteristics rather than on a representation interpreted in terms of learned stylistic norms. The particular piece, O. Messiaen's *Mode de valeurs et d'intensités*, was chosen because it is a very homogeneous and highly structured piece. It was written in 1949 in an atonal style without a tonal center, traditional scale structure, or conventional harmonies.

Historically, this piece was influential in the development of total serialism in the compositions of P. Boulez, K. Stockhausen, and others, because the values of pitch, duration, dynamics (loudness), and articulation are regulated according to strict principles. Because of these constraints, to be described below, selecting this piece allowed an additional psychological issue to be examined, namely, the independence of the dimensions of musical sound. Both theoretically and empirically, it has been traditional to treat separately the various dimensions of music, particularly duration and pitch. For example, the-

oretical treatments of harmony and counterpoint are only peripherally concerned with meter and rhythm. Summaries of psychological studies, such as those by Dowling and Harwood (1986) and Handel (1989), have separate chapters covering pitch and rhythm. Although it may be logically appealing to treat musical pitch and time independently, these dimensions may interact in complex ways in perception and memory.

Empirical results on the interactions between pitch and rhythmic structures are not extensive, and those that exist reach somewhat different conclusions concerning the psychological independence of these dimensions, depending on the experimental task. Palmer and Krumhansl (1987a, 1987b), for example, found temporal and pitch patterns have independent and additive influences on judged phrase structure. Monahan and Carterette (1985) found two kinds of dimensions underlying judgments of melodic similarity: rhythmic and tonal. Intersubject tradeoffs between the weights for the two kinds of dimensions indicate some degree of independence. Other results, however, suggest interactions between tonal and rhythmic patterns. Deutsch (1980), Boltz and Jones (1986), and Monahan, Kendall, and Carterette (1987) found memory interference when temporal and pitch groupings were set into conflict (see, however, Smith & Cuddy, 1989). Jones, Summerell, and Marshburn (1987) found that melodies presented in different rhythms were difficult to recognize. Thus, it seems clear that the two dimensions interact at some levels of perception. The present study approaches this issue in another way, namely, by independently varying pitch from the other dimensions of music and assessing the effect on recognition memory.

The underlying structure of the piece is shown schematically in Table 1. It is a piano piece, written in three separate lines, which will be called I (high), II (medium), and III (low). Within each line, all 12 pitches of the chromatic scale (C, C#, D, D#, E, etc.) occur, but two tones are never sounded simultaneously. Table 1 shows how the values of the musical dimensions (chroma, octave, duration, and dynamics) are coupled. *Duration* refers to the interval of time between the onset of a tone and the onset of the next tone. The composer also specifies 12 types of attack (articulations), each used once in each of the three lines. In the score, these are sometimes written into the music as durations (e.g., a staccato eighth note is rewritten as a sixteenth note and a sixteenth rest). However, the 12 different attack types did not appear to be performed in a consistent way by the recording pianist and so will not be considered further. The piece is constructed such that whenever a particular chromatic scale tone (called *chroma*) is sounded in a line, it is always sounded in a particular octave (numbered such that 4 means the octave beginning with middle C), for a particular duration (indicated in terms of quarter-note beats), and at a particular dynamic (loudness) level. As can be seen in Table 1, for example, whenever a B is sounded in line I,

Table 1
Specific Couplings between Chroma, Octave, Duration, and Dynamics
in O. Messiaen's *Mode de Valeurs et d'Intensités*

| Line I (top) | | | | Line II (middle) | | | | Line III (bottom) | | | |
|--------------|-----|-----------------|-----|------------------|-----|-----------------|-----|-------------------|-----|-----------------|-----|
| Chr | Oct | Dur | Dyn | Chr | Oct | Dur | Dyn | Chr | Oct | Dur | Dyn |
| E \flat | 7 | $\frac{1}{4}$ | ppp | G | 5 | $\frac{1}{2}$ | ff | E \flat | 5 | $\frac{1}{2}$ | ff |
| D | 7 | $\frac{1}{4}$ | ppp | C | 5 | $\frac{1}{2}$ | mf | D | 5 | 1 | ff |
| A | 6 | $\frac{3}{8}$ | ff | B \flat | 4 | $\frac{3}{8}$ | mf | A | 4 | 1 $\frac{1}{2}$ | mf |
| A \flat | 6 | $\frac{1}{2}$ | f | A \flat | 4 | 1 | p | G | 4 | 2 | pp |
| G | 6 | $\frac{3}{8}$ | mf | F | 4 | 1 $\frac{1}{4}$ | pp | F \sharp | 4 | 2 $\frac{1}{2}$ | p |
| F \sharp | 6 | $\frac{3}{8}$ | ff | E | 4 | 1 $\frac{1}{2}$ | p | C | 4 | 3 | f |
| E | 6 | $\frac{3}{8}$ | f | E \flat | 4 | 1 $\frac{3}{8}$ | p | A \flat | 3 | 3 $\frac{1}{2}$ | ff |
| C \sharp | 6 | 1 | mf | D | 4 | 2 | p | F | 3 | 4 | mf |
| C | 6 | 1 $\frac{1}{8}$ | ff | C3 | 4 | 2 $\frac{1}{4}$ | f | B | 2 | 4 $\frac{1}{2}$ | ff |
| B \flat | 5 | 1 $\frac{1}{4}$ | pp | B | 3 | 2 $\frac{1}{2}$ | f | E | 2 | 5 | ff |
| F | 5 | 1 $\frac{3}{8}$ | ff | F \sharp | 3 | 2 $\frac{3}{8}$ | f | B \flat | 1 | 5 $\frac{1}{2}$ | fff |
| B | 4 | 1 $\frac{1}{2}$ | p | F | 2 | 3 | f | C3 | 1 | 6 | fff |

Note—Chr = chroma, Oct = octave, Dur = duration (in beats), Dyn = dynamics, p = piano, f = forte, and m = mezzo.

it is always in the octave above middle C, its duration is always 1 $\frac{1}{2}$ beats, and it is always played at a ‘piano’ dynamic (denoted ‘p’).

In addition to these specific couplings, other, more general, correlations hold between the values on the musical dimensions. Within each line, higher tones are shorter; this also tends to be true across the three lines. In addition, pitch height correlates roughly with dynamics, with higher tones tending to be softer than lower tones. Finally, intervals in the higher pitch range tend to be smaller than intervals in the lower pitch range. The range of tones is approximately 2 $\frac{1}{2}$ octaves in the top line (I), 3 octaves in the middle line (II), and over 4 octaves in the bottom line (III). This also means, given the correlation between pitch height and duration, that the larger, lower intervals in the bass tend to be sounded at a slower rate than do the smaller intervals in the top line. Figure 1 shows the initial segment of the piece.

To assess recognition memory for this piece, it was necessary to construct some ‘foils’ to be used in the experiment. For this purpose, a variety of transformations were applied to the music; these are described schematically in Figure 2. Each of the transformations was motivated by certain musical considerations and/or prior experimental findings. Moreover, they preserved the general character of the piece. The rhythmic framework was maintained, leaving the tempo, density of tones, and occurrence of simultaneities fixed. The couplings between duration and dynamics (and attacks) were as in the original music, as was the overall pitch range. The transformations independently varied pitch from the other musical parameters in various ways in order to assess whether listeners remember either the precise couplings or the general correlations between pitch (chroma and octave) and the other musical parameters. The transformations were carried out according to systematic rules, but because of the complex, naturalistic materials, it was not possible to change one property without simultaneously changing others. For example, changing melodic inter-

vals (successive tones) in one line necessarily changes harmonic intervals (simultaneous tones) between lines.

In the first transformation (‘6 chromas in I’), 6 chromas were changed in the top line (I), each by 1 semitone (up or down a half step). For example, whenever a D7 was sounded in the original music, it was replaced by a C \sharp 7 in the transformed version; whenever an A6 was sounded in the original music, it was replaced by a B \sharp 6; and so on. The primary reason for choosing this transformation was that the contour (the pattern of rising and falling pitch) in the melodic figures in the top line is unchanged. Numerous studies have demonstrated that contour can be remembered even when precise interval information is not (e.g., Dowling, 1978; Dowling & Fujitani, 1971). The chroma changes were such that all 12 chromatic scale tones are sounded in the transformed version (as in the original music). This transformation also retains the pitch contour in the lower two lines and the harmonic intervals between them. This transformation, however, breaks the specific couplings between chroma and the other musical parameters. It also changes the melodic intervals within line I and the harmonic intervals between line I and lines II and III.

The second transformation (‘chromas I \leftrightarrow III’) interchanges the chromas in lines I and III. The lowest tone in line I (B) is replaced everywhere with the lowest tone in line III (C \sharp), and vice versa; the next lowest tone in line I (F) is replaced everywhere with the next lowest tone in line III (B \flat) and vice versa; and so on. This is done in order to match as well as possible the top and bottom tones of the entire pitch range in the original and transformed versions. As with the first transformation, this transformation breaks the specific couplings between chroma and the other musical parameters and changes specific intervals both within and between lines. In addition, intervals at the top of the overall pitch range tend to be larger than in the original music, whereas intervals at the bottom of the overall pitch range tend to be smaller. Thus, this transformation violates the general tendency

Modéré

Figure 1. The initial segment of Messiaen's *Mode de valeurs et d'intensités*. Copyright 1949 Durand S.A. Reproduced by permission of the publisher. Sole representative U.S.A. Theodore Presser Company.

in music for smaller intervals to appear in melodies and in harmonies in the higher pitch range than in the lower pitch range (Dowling & Harwood, 1986; Hutchinson & Knopoff, 1979). All 12 chromas are sounded in each line, however, and the pitch contour of melodic figures is unchanged. Notice also that the top 3 chromas of the top line (which may be salient because they are the highest tones in the piece) are unchanged, and the next two lower tones are changed by only a semitone. The analogous property also holds, of course, for the lower line.

The third transformation ("inversion midrange") inverts each of the three lines around the midpoint of its pitch range. To do this, the tone halfway between the top and the bottom tone of each line was determined, and then all intervals were reflected around this midpoint. For example, the midpoint of the top line is C#6; whenever a tone in the original music was sounded above this midpoint by a certain number of semitones, it was replaced with a tone that many tones below the midpoint, and vice versa.

Inversions of this sort appear in music and take a central role in serial (dodecaphonic) music. Francès (1988), Dowling (1972), and Krumhansl, Sandell, and Sergeant (1987) obtained evidence that listeners perceive some degree of melodic invariance under inversions of this sort. This transformation is such that it preserves the property that all 12 chromas are represented in each line. However, few other properties of the original music remain unchanged. Perhaps most importantly, the pitch contour of melodic figures is inverted. In addition, specific couplings between chroma, octave, duration, and dynamics are broken, and harmonic intervals between lines are changed. Also, within lines, lower tones are now of shorter duration than are higher tones. Also, downward intervals predominate in the original music (as in music generally; Vos & Troost, 1989), whereas upward intervals predominate in this transformation. The global correlation across lines between pitch height and duration, however, is maintained.

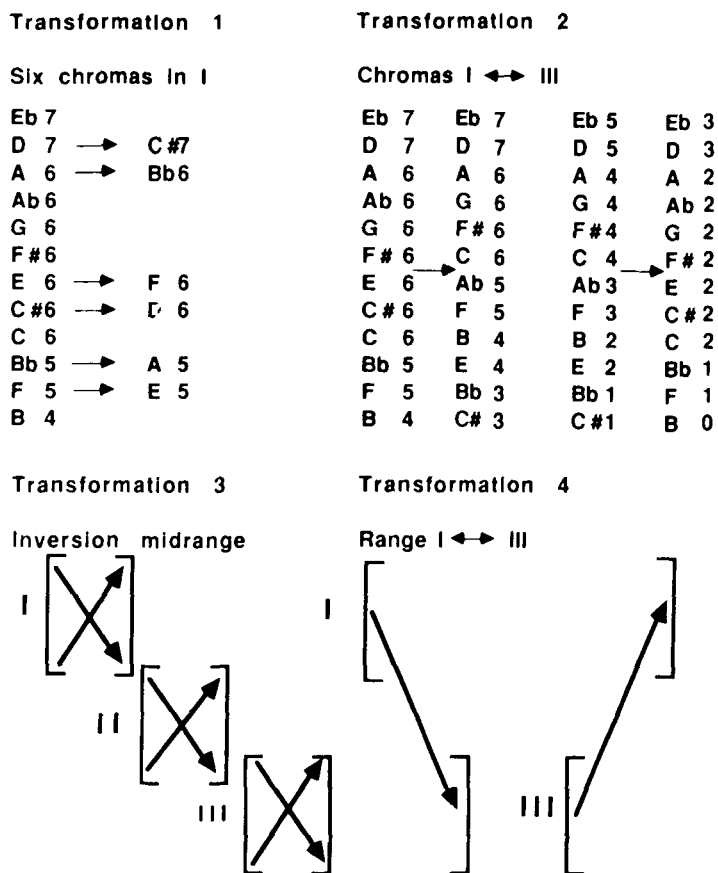


Figure 2. The four transformations that were applied to the music. Transformation 1 changes 6 chromas in the top line, each by one step only. Transformation 2 interchanges the chromas in the top and bottom lines. Transformation 3 inverts each of the three lines around its midpoint. Transformation 4 shifts the top line to the range of the bottom line, and vice versa.

The fourth, and final, transformation ("range I ↔ III") simply moves the top line down to the range of the bottom line, and vice versa. This was done such that the original and transformed versions share the highest tone in line I and the lowest tone in line III, keeping the overall pitch range the same. It is fairly common in music for a melodic figure to pass back and forth between the higher range and the lower range, as between instruments in a string quartet or between hands on the piano. The primary effect of this transformation is to break the global correlation between pitch height and duration, as well as the rough correlation between pitch height and dynamics and between pitch height and interval size. In addition, intervals formed by tones in different lines are altered. However, within lines, contour and melodic intervals are unchanged, and the correlation between pitch height and duration is maintained.

In the experiments, listeners heard just the first half of the piece, which is slightly longer than 2 min. This section is indicated by the hatched segment of Figure 3.

There were six types of test segments. The first segment type was an excerpt from the part of the piece listeners had heard; their positions in the piece are indicated by the rectangles under the label "original old." The second was an excerpt from the part of the piece that they had not heard; their positions in the piece are indicated by the rectangles under the label "original new." The remaining four segment types were the four transformations just described of the "original new" segments. The listeners' task was to judge how likely it was that each excerpt came from the piece, whether or not it came from the part of the piece they had heard. In other words, a correct "yes" answer would be given to the first two kinds of test segments. Listeners were told that the piece was very homogeneous throughout. After hearing the first half of the piece and six test segments (one of each type, randomly selected), listeners heard the first half of the piece again and six more test segments (again, one of each type, randomly selected without replacement). This process was repeated a total of six times. Thus, the design

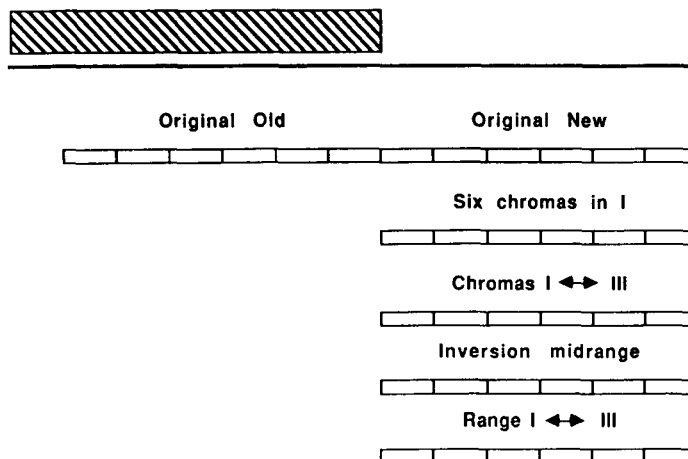


Figure 3. The materials used in the study. The listeners heard just the first half of the piece, indicated by the hatched segment. Six types of test trials were presented: original old (from the part of the piece listeners heard), original new (from the part of the piece listeners did not hear), and each of the four transformations as described in Figure 2.

of the experiments made it possible to assess the effect of repeated hearings of the piece without ever repeating a test segment.

If original old segments are recognized accurately, this indicates that listeners have detailed memory for the musical surface. If original new segments also are recognized accurately, this indicates that listeners are able to abstract the surface characteristics of the piece from the part they had heard and generalize to the rest of the piece. In addition, performance on the transformed versions may be informative about the kinds of surface regularities to which listeners are sensitive. The four transformations represent a variety of systematic alterations of the underlying structure of the piece, each with a number of different consequences for melodic and harmonic intervals, contour, interval size, and other factors as have been described. Despite the multiple structural changes resulting from each of the different transformations, comparing the results for them may reveal something about the kinds of surface regularities to which listeners are sensitive. The primary objective of the study, however, was to assess the accuracy of recognition memory for music in an unfamiliar style for which listeners would not have schematic knowledge to use in interpreting and remembering the music.

EXPERIMENT 1

Method

Apparatus and Stimulus Materials. The stimulus materials were based on a performance of O. Messiaen's *Mode de valeurs et d'intensités* by Pierre-Laurent Aimard of the Ensemble InterContemporain (Pierre Boulez, director). The performance was played on a Yamaha KX-88 keyboard with a piano timbre produced by Yamaha TX-816 and DX-7 synthesizers, amplified and played over loudspeakers. A Macintosh Plus computer (with MIDI interface) recorded the timing (onset and duration) and velocity of each keypress using the *Performer* software. Two complete recordings

of the piece were made; the performance that was judged more satisfactory by the pianist served as the basis for the materials used in the experiment. Its total duration was 4.29 min. The *Performer* software was also used to create the transformed versions of the piece and to extract the excerpts.

The stimuli were played back during the experiment using a Macintosh Plus computer, MIDI interface, and an Akai S-900 Digital Sampler (producing a piano timbre); the analog output was amplified and played at a comfortable listening level over loudspeakers. The section of the piece played before each block of trials consisted of the first 2.22 min. The test segments were each 19.07 sec in duration. The six types of test segments were constructed as described earlier. The original old segments did not include the first 19.07 sec of the piece, and the original new segments did not include the last 9.54 sec of the piece (because these parts of the piece might be especially well remembered). The transformations were based on materials of the second half of the piece.

Subjects. The 25 listeners who participated in the experiment were each paid £3.00. They were all music students at City University, London. On average, they had received 16.46 years tuition on various musical instruments. Fourteen listeners were first-year students, 10 were second-year students, and 1 was a third-year student. Three were familiar with the piece prior to the experiment: one had heard it, and two had studied it analytically.

Procedure. The listeners were told that the experiment investigated the ability to judge whether or not a musical segment belongs to a larger piece of music of which they would hear only the first half. They were also told that the piece is quite homogeneous from beginning to end and that the test segments (1) may come from the first half of the piece (which they would hear), (2) may come from the second half of the piece (which they would not hear), or (3) may be transformations of material from the second half of the piece. The listeners' task was to indicate, on an 8-point scale, how sure they were that the segment they heard belonged to the piece (1 = *definitely does not belong to the piece*; 8 = *definitely does belong to the piece*). It was emphasized that "belongs to the piece" means the test segment came from either the first half of the piece (which they heard) or the second half (with they did not hear).

The experimental session began with the first half of the piece played once, followed by three practice trials. This was followed by six blocks of experimental trials, each of which began with the playing of the first half of the piece. Each of the six types of test

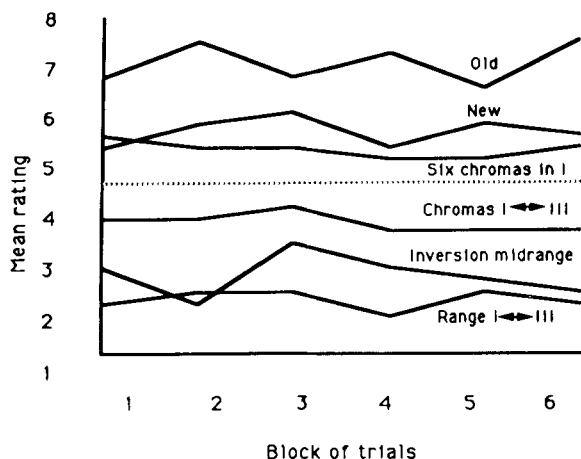


Figure 4. The average responses for each test segment type for each of the six blocks of trials in Experiment 1. The listeners were music students, most of whom were unfamiliar with the piece before the experiment.

segment was represented once in each block of trials in random order. No test segments were repeated during the experiment. The subjects were run in groups, with different random orders of test segments for different groups.

Results and Discussion

The average rating for each segment type is shown in Figure 4 for the six blocks of experimental trials. Recall that the listeners heard the first half of the piece once before the practice trials and once again before each block of six trials. The most striking result is that the ratings did not change with repeated hearings. The main effect of block number was not significant [$F(5,120) = 1.298$, n.s.], nor was the interaction between block number and test segment type [$F(25,600) = 1.298$, n.s.]. Thus, additional exposure to the piece did not result in greater differentiation among the segment types, suggesting the basis for judging whether or not the test segments come from the piece develops rapidly. Because block number had no effect and did not interact with test segment type, the remaining analyses were performed on the data averaged over block number; these values are shown at the top of Table 2.

Test segment type had a strong effect [$F(5,120) = 101.136$, $p < .0001$]. To see whether the different excerpt types were reliably recognized or rejected, the data averaged across blocks were compared with the neutral value on the rating scale of 4.5 (halfway between the extremes of the response range). The listeners did not show strong biases; the average of the ratings was 4.47, which was not statistically different from 4.5 [$t(24) = -.2806$, $p = .782$]. Highest ratings went to original old segments, which were very reliably judged as having come from the piece [the t value comparing the ratings to the neutral level of 4.5 was $t(24) = 18.058$, $p < .0001$]. This result shows excellent recognition of segments drawn from the part of the piece the listeners had heard. The listeners were also very accurate in judging the original new segments as hav-

ing come from the piece [$t(24) = 6.645$, $p < .0001$]. Thus, listeners can abstract the surface characteristics from hearing the first half of the piece and use this as a basis for correctly judging segments from the second half of the piece, which they had not heard. The difference between original old and original new test segments was significant (by a Scheffé post hoc test for differences between means), indicating that memory for the specific passage was having an effect over and above a general sense of the piece's characteristics.

However, the listeners also gave high ratings to the transformation in which six chromas in the top line were changed (6 chromas in I). The ratings for these excerpts were higher than the neutral point on the response scale [$t(24) = 3.267$, $p = .0033$]. In fact, a Scheffé post hoc test showed the ratings for these test segments did not differ from those for the original new segments. This transformation breaks the specific couplings between chroma and the other musical parameters and slightly alters the sizes of pitch intervals both within the top line and between the top line and the middle and lower lines. Apparently, these changes were not salient enough to enable the listeners to reject these segments. It seems likely that the properties that are invariant under this transformation (particularly pitch contour and the general correlations within and across lines between pitch height, duration, interval size, and dynamics) engendered a strong sense of similarity to the original music.

The remaining three transformation types were reliably rejected as not having come from the piece. The transformation interchanging the chromas in the top and bottom lines (Chromas I ↔ III) received ratings lower than the neutral level of 4.5 [$t(24) = -2.552$, $p = .0175$]. The greater number of interval changes in both top and bottom lines (as well as changes in intervals between lines) probably contributed to the better differentiation of this transformation than the transformation that changed just 6 chromas in the top line. In addition, this transformation placed larger intervals in the top line and smaller intervals in the bottom lines, relative to the original music.

Table 2
Average Ratings for Six Test Segment Types in Experiment 1

| Segment Type | Average Rating | Standard Error |
|---|----------------|----------------|
| All Subjects | | |
| Original Old | 7.02 | .14 |
| Original New | 5.74 | .19 |
| 6 Chromas in I | 5.25 | .23 |
| Chromas I ↔ III | 3.87 | .25 |
| Inversion Midrange | 2.70 | .21 |
| Range I ↔ III | 2.25 | .17 |
| Subjects Previously Unfamiliar With Piece | | |
| Original Old | 7.04 | .15 |
| Original New | 5.81 | .20 |
| Six Chromas in I | 5.42 | .22 |
| Chromas I ↔ III | 3.94 | .28 |
| Inversion Midrange | 2.75 | .23 |
| Range I ↔ III | 2.25 | .18 |

Very low ratings went to the transformation inverting each of the three lines around its midpoint (inversion midrange). These segments were judged reliably as not having come from the piece [$t(24) = -8.76, p < .0001$]. The most salient feature distinguishing these segments from the original is the inversion of contour and the reversed correlation between pitch height and duration within lines. In addition, ascending intervals predominate in this transformation, whereas descending intervals predominate in the original. Very low ratings also went to the transformation shifting the top line to the pitch range of the bottom line, and vice versa (Range I \leftrightarrow III). These segments, very reliably rejected [$t(24) = -13.473, p < .0001$], most obviously broke the global correlation across lines between pitch height and duration; this transformation contained many low, fast tones, relative to the original music. These last two conditions were not reliably different from one another.

To see whether familiarity with the piece prior to the experiment had an effect, the ratings of the 3 listeners who knew the piece previously were excluded from the analysis. The average ratings for the remaining 22 listeners are shown at the bottom of Table 2. The results were virtually identical, and statistical analyses found the same effects to be reliable. The only exception to this is for the transformation interchanging the chromas in the top and bottom lines (Chromas I \leftrightarrow III). These again received ratings lower than the neutral level of 4.5, but the effect was only marginally significant [$t(21) = -2.020, p = .0563$].

In summary, these results show very accurate memory for the part of the piece presented to the listeners and an ability to generalize to the remainder of the piece. Some of the factors underlying this ability are revealed by the pattern of responses on the transformations. Transformations of the material were easily rejected if they violated the general correlation between pitch height, duration, dynamics, and interval size, and/or inverted the contour of melodic figures. These properties appear to be intrinsic to the sense of surface characteristics abstracted from the piece. There is little evidence, however, that precise pitch intervals or the specific couplings (shown in Table 1) between chroma and the other musical parameters (octave, duration, and dynamics) were accurately remembered.

EXPERIMENT 2

The results of Experiment 1 showed that, although the listeners quite rapidly developed a sense of the surface characteristics of the piece, they apparently did not base their judgments of whether segments did or did not belong to the piece on the specific couplings between chroma and the other musical parameters (octave, duration, and dynamics). This raises the question as to whether the listeners could not code these specific couplings (because, for example, they were processed independently) or simply did not consider the possibility of their existence. This kind of compositional constraint is rare, and the listeners may not have attended to specific pairings when listen-

ing to the piece. Alternatively, or in addition, the listeners may not have had enough musical training to encode the pitch information in relation to the other parameters. To examine this question, the same procedure was employed with a group of highly trained, professional musicians who were explicitly instructed about the structure of the piece before the experiment began. If their performance is similar, then neither lack of knowledge about the structure of the piece nor an inability to code couplings between chroma and the other parameters would account for the results of Experiment 1.

Methods

Apparatus and Stimulus Materials. The apparatus and stimulus materials were the same as those used in Experiment 1, except the stimuli were played back during the experiment using a Yamaha TX-802 synthesizer that produced a piano timbre.

Subjects. Nine professional musicians, mostly composers at the Institut de Recherche et Coordination Acoustique/Musique, Paris, volunteered to participate in Experiment 2. Six subjects knew the piece previously; 1 had studied it analytically.

Procedure. The procedure was the same as that used in Experiment 1, with the exception that, prior to the experiment, the structure of the piece was explained. Essentially, this instruction consisted of showing the listeners a chart containing the same information as is presented in Table 1.

Results

Figure 5 shows the results for the professional musicians who were informed of the structure of the piece prior to the experiment. As can be seen, their results were very similar to those for the music students, most of whom were unfamiliar with the piece. Again, repeated hearings of the music had little effect on the results. Block number did not have a significant main effect [$F(5,40) = 1.686, n.s.$], nor did it interact with test segment type [$F(25,200) = 1.009, n.s.$].

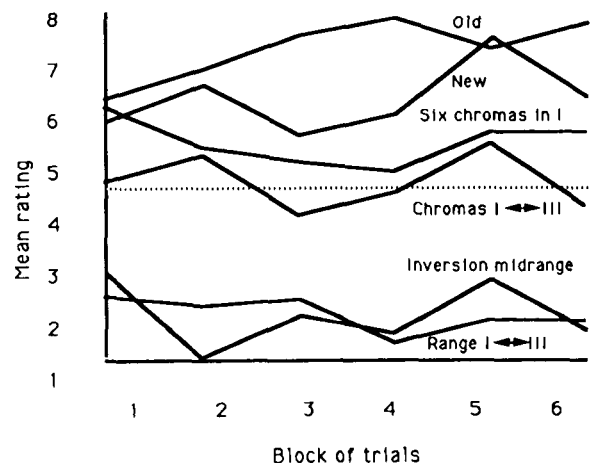


Figure 5. The average responses for each test segment type for each of the six blocks of trials in Experiment 2. The listeners were professional musicians, who were informed about the structure of the piece before the experiment.

Table 3
Average Ratings of Professional Musicians
for Six Segment Types in Experiment 2

| Segment Type | Average Rating | Standard Error |
|--------------------|----------------|----------------|
| Original Old | 7.33 | .23 |
| Original New | 6.32 | .19 |
| 6 Chromas in I | 5.50 | .34 |
| Chromas I → III | 4.65 | .32 |
| Inversion Midrange | 1.91 | .24 |
| Range I → III | 1.96 | .36 |

The analysis of the data averaged over blocks, shown in Table 3, revealed a very strong effect of test segment type [$F(5,40) = 83.438, p < .0001$]. The ordering of the conditions was essentially the same as before. These listeners also did not show a strong bias in the use of the response scale; the average rating was 4.61, not statistically different from 4.5 [$t(24) = .633, p = .545$]. Highest ratings were given to original old segments, which were judged as coming from the piece with high accuracy [the t value comparing the ratings with the neutral level of 4.5 was $t(8) = 12.369, p < .0001$]. Next highest ratings went to original new segments, which the listeners rated significantly above the neutral point on the response scale [$t(8) = 9.587, p < .0001$].

Again, there were frequent false alarms to the transformation in which 6 chromas in the top line were changed (6 chromas in I); these were rated higher than the neutral point on the scale [$t(8) = 2.949, p = .0184$]. These listeners also had difficulty with the transformation exchanging chromas between the top and bottom lines (Chromas I → III); ratings for these were not significantly different from the neutral point on the response scale [$t(8) = 0.471, p = .6505$]. However, the final two transformations were consistently rejected [$t(8) = -10.991, p < .0001$, and $t(8) = -6.989, p < .0001$, for inversion midrange and Range I → III transformations, respectively]. The Scheffé post hoc test for differences between means showed the differences between the last two conditions to be the only nonsignificant difference between conditions.

To summarize, the results for Experiments 1 and 2 were very similar. This similarity suggests that the highly trained musicians, who were informed of the structure of the piece, based their responses on criteria similar to that of the less trained musicians, who were unfamiliar with the piece. The professional musicians made somewhat sharper distinctions between the segments actually drawn from the piece and the more obvious transformations, but they actually performed somewhat worse (although not significantly so) than did the music students on one of the transformations (in which the chromas in the top and bottom lines were interchanged). When asked about this, the listeners in Experiment 2 rationalized their performance by saying this seemed like the kind of transformation Messiaen might have employed. In addition, they might have been more sensitive to the fact that, in lines I and III, the original and transformed versions share a number of chromas (although sometimes in different octave).

SUMMARY AND CONCLUSIONS

The primary result of these two experiments was the remarkable accuracy with which the listeners recognized short excerpts drawn from the section of the piece they heard during the experiments. Accuracy levels were high in the first block of test trials (after the subjects had heard the passage twice) and did not increase with repeated hearings. This was true for both groups of subjects: the undergraduate music students (most of whom were unfamiliar with the piece prior to the experiment) and the professional musicians (most of whom knew the piece in advance and all of whom were instructed as to its structure). This ability is surprising in light of two aspects of the musical materials used in the experiments. First, the test excerpts were short and were excised from the music without regard for possible natural subdivisions of the music. Second, the compositional method of the music is highly unusual. The compositional constraints regulating the choice of pitch, duration, and dynamic values prevent the music from conforming to conventional scalar, harmonic, metrical, or rhythmic patterns. Thus, knowledge of these conventional patterns could not be used to help interpret and remember the sounded events.

It would seem, then, that listeners have the capacity to remember large amounts of surface detail in music. In this respect, memory for music might be more similar to memory for pictures than to memory for language. Even though, as summarized in the introduction, listeners can remember verbatim information for language, it appears that long-term memory for language consists largely of abstract meaning or gist. In contrast, the memory capacity for surface or perceptual aspects of pictures is large. Shepard (1967) presented 612 pictures of common scenes, which were later recognized at a rate of 97% correct in a forced-choice recognition task. In similar experiments, Standing, Conezio, and Haber (1970) found 91% accuracy with 2,500 pictures, and Standing (1973) found 73% accuracy with 10,000 pictures. In a study by Brown and Scott (1971), 4-year-old children were also quite accurate (98% correct on a set of 100 pictures). Recognition rates in these studies decreased somewhat with time but remained well above chance at extended delays. For example, Shepard's (1967) subjects were performing at 58% correct after 4 months. Thus, it appears that long-term memory capacity for surface or perceptual information is great for both pictures and music. That research on memory for music tends to emphasize memory limitations may stem from the fact that the majority of studies use short, artificially constructed sequences designed to isolate the structural variable(s) of interest.

The second result was that the listeners in the present experiments could abstract the surface characteristics from the part of the piece they had heard and apply it to the rest of the piece. Segments from the rest of the piece were quite accurately judged as having come from the piece. Again, performance did not improve with repeated hearings. Thus, in addition to forming specific, detailed sur-

face memories, the listeners apparently learned surface characteristics that could be used to judge the new segments. A similar finding was obtained in a study of children's sensitivity to style (Gardner, 1973). Children, ranging from ages 6 to 19, were presented with pairs of short musical excerpts (15 sec in duration) from pieces written in Baroque, Classical, Romantic, and Modern periods. Within a pair, the two excerpts were either from the same piece or from different pieces (which were written by different composers and sometimes came from different periods). All groups of subjects performed well above chance, although performance improved significantly with age. Excerpts coming from different periods were somewhat easier to identify as coming from different pieces, especially for the older subjects. Both this and the present study, then, show an ability to perceive the similarity of two excerpts from the same piece. In Gardner's (1973) study, however, the fact that "different" pairs came from different pieces probably meant they were more distinctive than were the "foils" in the present experiments. Indeed, the listeners in this study had difficulty rejecting some of the transformed versions.

The final results come from performance on the transformed versions of the music. These were not always consistently rejected, suggesting certain limits in the precision with which the piece's surface characteristics are abstracted. This particular piece is constructed with a unique set of constraints that couple together values of pitch (chroma and octave), duration, and dynamics. One of the transformations that broke these specific couplings was difficult for the listeners to detect, suggesting that these musical parameters are encoded independently at some level. This was true even for the highly trained listeners, who had been informed about these couplings in advance of the experiment. Although this transformation altered the majority of melodic intervals in the top line (and many of the harmonic intervals between lines), it maintained the melodic contour (the pattern of rising and falling pitch). In contrast, the transformation that inverted contour (by inverting each line around its midrange) was easily rejected. This result fits with numerous studies (e.g., Divenyi & Hirsh, 1974; Dowling, 1978; Dowling & Fujitani, 1981) showing that contour information is relatively easy to encode.

Finally, the compositional constraints governing the piece entail a number of rough correlations between dimensions. High tones tend to be shorter, softer, and closer together in pitch distance. The results suggest that listeners are sensitive to these rough correlations in the piece. Transformations reversing them were easier to reject than are transformations that maintain them. The fact that music generally has smaller intervals in the higher pitch range than in the lower pitch range may also have aided the listeners in rejecting certain of the transformations. In a similar vein, the predominance of descending intervals in music generally (as in the present piece) may have affected judgments. Thus, the influence of extraopus regularities cannot be ruled out absolutely.

In sum, the results of the present study show that musically trained listeners have surprisingly accurate recognition memory for the musical surface. This result stands in contrast to many studies showing frequent errors in recognition memory for music (e.g., Bharucha & Krumhansl, 1983; Cuddy et al., 1979; Dowling, 1978). These studies, however, are cases in which the artificially constructed materials would be expected to be assimilated to learned compositional norms. It will be of interest, then, to study surface memory as a function of musical style and training in future studies. In contrast to the accurate recognition memory, the results of the present study suggest that the process of abstracting and generalizing to novel segments is rather imprecise. Although the listeners exhibited the ability to generalize to the part of the piece they had not heard, recognizing new segments reliably, they also accepted certain transformations preserving some of the general surface characteristics. Thus, somewhat looser criteria may underlie this abstraction process. This flexibility would serve well the identification of similarity, for example, between theme and variations (Pollard-Gott, 1983; Welker, 1982) or pieces with a common underlying form (Rosner & Meyer, 1982).

REFERENCES

- ANDERSON, J. R. (1974). Verbatim and propositional representations of sentences in immediate and long-term memory. *Journal of Verbal Learning & Verbal Behavior*, *13*, 149-162.
- ANDERSON, J. R., & PAULSON, R. (1977). Representation and retention of verbatim information. *Journal of Verbal Learning & Verbal Behavior*, *16*, 439-451.
- BARTLETT, J. C., & DOWLING, W. J. (1988). Scale structure and similarity of melodies. *Journal of Experimental Psychology: Human Perception & Performance*, *6*, 501-515.
- BATES, E., KINTSCH, W., FLETCHER, C. R., & GIULIANI, V. (1980). The role of pronominalization and ellipsis in texts: Some memory experiments. *Journal of Experimental Psychology: Human Learning & Memory*, *6*, 676-691.
- BATES, E., MASLING, M., & KINTSCH, W. (1978). Recognition memory for aspects of dialogue. *Journal of Experimental Psychology: Human Learning & Memory*, *4*, 187-197.
- BHARUCHA, J. J., & KRUMHANSL, C. L. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition*, *13*, 63-102.
- BHARUCHA, J. J., & PRYOR, J. H. (1986). Disrupting the isochrony underlying rhythm: An asymmetry in discrimination. *Perception & Psychophysics*, *40*, 137-141.
- BOLTZ, M., & JONES, M. R. (1986). Does rule recursion make melodies easier to reproduce? If not, what does? *Cognitive Psychology*, *18*, 389-431.
- BRANSFORD, J. D., BARCLAY, J. R., & FRANKS, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, *3*, 193-209.
- BROWN, A. L., & SCOTT, M. S. (1971). Recognition memory for pictures in preschool children. *Journal of Experimental Child Psychology*, *11*, 401-412.
- CUDDY, L. L., COHEN, A. J., & MILLER, J. (1979). Melody recognition: The experimental application of musical rules. *Canadian Journal of Psychology*, *33*, 148-157.
- DEUTSCH, D. (1980). The processing of structured and unstructured tonal sequences. *Perception & Psychophysics*, *28*, 381-389.
- DIVENYI, P. L., & HIRSH, I. J. (1974). Identification of temporal order in three-tone sequences. *Journal of the Acoustical Society of America*, *56*, 144-151.

- DOWLING, W. J. (1972). Recognition of melodic transformations: Inversion, retrograde, and retrograde inversion. *Perception & Psychophysics*, **12**, 417-421.
- DOWLING, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, **85**, 341-354.
- 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., & FUJITANI, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America*, **49**, 524-531.
- DOWLING, W. J., & HARWOOD, D. L. (1986). *Music cognition*. Orlando, FL: Academic Press.
- FRANCÉS, R. (1988). *The perception of music* (W. J. Dowling, Trans.). Hillsdale, NJ: Erlbaum. (Originally published, 1958)
- GARDNER, H. (1973). Children's sensitivity to musical styles. *Merrill-Palmer Quarterly*, **19**, 67-77.
- GARROD, S., & TRABASSO, T. (1973). A dual-memory information processing interpretation of sentence comprehension. *Journal of Verbal Learning & Verbal Behavior*, **12**, 155-167.
- GRAESSER, A., & MANDLER, G. (1975). Recognition memory for the meaning and surface structure of sentences. *Journal of Experimental Psychology: Human Learning & Memory*, **104**, 238-248.
- HANDEL, S. (1989). *Listening: An introduction to the perception of auditory events*. Cambridge, MA: MIT Press.
- HUTCHINSON, W., & KNOPOFF, L. (1979). The significance of the acoustic component of consonance in Western music. *Journal of Musicological Research*, **3**, 5-22.
- JAMES, C. T., THOMPSON, J. G., & BALDWIN, J. M. (1973). The reconstructive process in semantic memory. *Journal of Verbal Learning & Verbal Behavior*, **12**, 51-63.
- JARVELLA, R. J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning & Verbal Behavior*, **10**, 409-416.
- JONES, M. R., SUMMERELL, L., & MARSHBURN, E. (1987). Recognizing melodies: A dynamic interpretation. *Quarterly Journal of Experimental Psychology*, **39A**, 89-121.
- KEENAN, J. M., MACWHINNEY, B., & MAYHEW, D. (1977). Pragmatics in memory: A study of natural conversation. *Journal of Verbal Learning & Verbal Behavior*, **16**, 549-560.
- KINTSCH, W., & BATES, E. (1977). Recognition memory for statements from a classroom lecture. *Journal of Experimental Psychology: Human Learning & Memory*, **3**, 150-159.
- KRUMHANSL, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology*, **11**, 346-374.
- KRUMHANSL, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.
- KRUMHANSL, C. L. (1991). Music psychology: Tonal structures in perception and memory. *Annual Review of Psychology*, **42**, 277-303.
- KRUMHANSL, C. L., & CASTELLANO, M. A. (1983). Dynamic processes in music perception. *Memory & Cognition*, **11**, 325-334.
- KRUMHANSL, C. L., SANDELL, G. J., & SERGEANT, D. C. (1987). The perception of tone hierarchies and mirror forms in twelve-tone serial music. *Music Perception*, **5**, 31-78.
- MONAHAN, C. B., & CARTERETTE, E. C. (1985). Pitch and duration as determinants of musical space. *Music Perception*, **3**, 1-32.
- MONAHAN, C. B., KENDALL, R. A., & CARTERETTE, E. C. (1987). The effect of melodic and temporal contour on recognition memory for pitch change. *Perception & Psychophysics*, **41**, 576-600.
- OFFIR, C. E. (1973). Recognition memory for presuppositions of relative clause sentences. *Journal of Verbal Learning & Verbal Behavior*, **12**, 636-643.
- PALMER, C., & KRUMHANSL, C. L. (1987a). Independent temporal and pitch structures in perception of musical phrases. *Journal of Experimental Psychology: Human Perception & Performance*, **13**, 116-126.
- PALMER, C., & KRUMHANSL, C. L. (1987b). Pitch and temporal contributions to musical phrase perception: Effects of harmony, performance timing, and familiarity. *Perception & Psychophysics*, **41**, 505-518.
- PALMER, C., & KRUMHANSL, C. L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 728-741.
- POLLARD-GOTT, L. (1983). Emergence of thematic concepts in repeated listening to music. *Cognitive Psychology*, **15**, 66-94.
- ROSNER, B. S., & MEYER, L. B. (1982). Melodic processes and the perception of music. In D. Deutsch (Ed.), *The psychology of music* (pp. 317-341). New York: Academic Press.
- SACHS, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, **2**, 437-442.
- SHEPARD, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning & Verbal Behavior*, **6**, 156-163.
- SHEPARD, R. N., & JORDAN, D. (1984). Auditory illusions demonstrating that tones are assimilated to an internalized musical scale. *Science*, **226**, 1333-1334.
- SMITH, K. C., & CUDDY, L. L. (1989). Effects of metric and harmonic rhythm on the detection of pitch alterations in melodic sequences. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 457-471.
- STANDING, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, **25**, 207-222.
- STANDING, L., CONEZIO, J., & HABER, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2560 visual stimuli. *Psychonomic Science*, **19**, 73-74.
- VOS, P. G., & TROOST, J. M. (1989). Ascending and descending melodic intervals: Statistical findings and their perceptual relevance. *Music Perception*, **6**, 383-396.
- WANNER, H. E. (1968). *On remembering, forgetting, and understanding sentences: A study of the deep structure hypothesis*. Unpublished doctoral dissertation, Harvard University, New Haven, CT.
- WELKER, R. L. (1982). Abstraction of themes from melodic variations. *Journal of Experimental Psychology: Human Perception & Performance*, **8**, 435-447.

(Manuscript received June 11, 1990;
revision accepted for publication January 30, 1991.)