The Perception of Tone Hierarchies and Mirror Forms in Twelve-Tone Serial Music
Author(s): Carol L. Krumhansl, Gregory J. Sandell and Desmond C. Sergeant
Source: *Music Perception: An Interdisciplinary Journal*, Vol. 5, No. 1, Organization of Pitch Structures (Fall, 1987), pp. 31-77
Published by: University of California Press
Stable URL: http://www.jstor.org/stable/40285385
Accessed: 03/04/2013 15:54

Music Perception
Fall 1987, Vol. 5, No. 1, 31–78

The Perception of Tone Hierarchies and Mirror Forms in Twelve-Tone Serial Music

CAROL L. KRUMHANSL & GREGORY J. SANDELL
Cornell University

DESMOND C. SERGEANT
Roehampton Institute

Four experiments are reported in which the materials are derived from two 12-tone serial compositions (Schoenberg's Wind Quintet and String Quartet, No. 4). Two experiments use the probe tone method (Krumhansl & Shepard, 1979) to assess factors contributing to tone prominence in serial music. The contexts in Experiment 1 are musically neutral statements of the complete or incomplete tone rows; the contexts in Experiment 4 are excerpts from the two pieces. Two experiments use a classification task to evaluate whether the prime form of the row is perceived as similar to its mirror forms (inversion, retrograde, and retrograde inversion). The materials are neutral presentations of the forms (Experiment 2) or excerpts from the pieces (Experiment 3). Large individual differences are found. A subgroup of listeners, with more music training on average, show the following effects in the probe tone experiments: low ratings for tones sounded more recently in the contexts and high ratings for tones not yet sounded; low ratings for tones fitting with local tonal implications; similar patterns for the neutral contexts and the musical excerpts. The remaining listeners show the opposite effects. Classification accuracy of mirror forms is above chance and is higher for the neutral sequences than the musical excerpts; performance is correlated with music training. The experiments show that some, but not all, listeners can perceive invariant structures in serial music despite mirror transformations, octave transpositions of tones, and variations of rhythm and phrasing.

THIS article describes a series of experiments addressing a number of perceptual issues that arise from dodecaphonic or twelve-tone serial music. The materials for the experiments were derived from two compositions by Arnold Schoenberg: the Wind Quartet, Op. 26, and the String

Requests for reprints should be addressed to: Carol L. Krumhansl, Department of Psychology, Uris Hall, Cornell University, Ithaca, NY 14853.

Quartet, No. 4, Op. 37. We begin with a brief description of the technique of serial composition, focusing on elementary characteristics exhibited by compositions in Schoenberg's third stylistic period, during which these two pieces were written. We rely primarily on Schoenberg's own theoretical treatment of the compositional method he introduced during this period. Following this introduction, we note briefly some characteristics of the two pieces from which the materials of the experiments were drawn, and finally give an overview of the experiments.

Schoenberg (1874–1951) is a central figure in the development of new processes in pitch structuring in Western twentieth-century music. His works are generally classified into three periods. The compositions in the first period, characterized by the use of expanded tonality, are clearly rooted in the traditions of Austrian–German music. The second period exhibits a radical shift from traditional concepts to a style in which pitch materials are freed from diatonic–functional associations. Constructed so as to prevent reference to tonal harmonic and melodic structures, music of this period is called "freely atonal." The third period introduces the technique of dodecaphony or twelve-tone serialism. This technique, which has had a profound effect on twentieth-century composition and theory, is described in Schoenberg's writings collected in *Style and Idea* and the footnotes and appendices added to the revised edition of his textbook *Harmonielehre* (*Theory of Harmony*) (1922/1978).

In his essay, "Composition with twelve tones" (1941/1975, p. 218), Schoenberg traces his thinking underlying the development of twelve-tone serialism:

> After many unsuccessful attempts during a period of approximately twelve years, I laid the foundations for a new procedure in musical construction which seemed fitted to replace those structural differentiations provided formerly by tonal harmonies. I called this procedure *Method of Composing with Twelve Tones Which are Related Only with One Another*. This method consists primarily of the constant and exclusive use of a set of twelve different tones. This means, of course, that no tone is repeated within the series and that it uses all twelve tones of the chromatic scale, though in a different order.

Each composition is based on a specified order of the 12 tones of the chromatic scale (which is different for each composition). This is called the basic set, series, or tone row. Both melodic and harmonic elements are derived from it, and Schoenberg (1941/1975) describes the technique in terms of the following analogies to tonal music.

> [The basic set] is invented to substitute for some of the unifying and formative advantages of scale and tonality. The scale [in tonal music] is the source of many figurations, parts of melodies and melodies themselves, ascending and descending passages, and even broken chords. In ap-

proximately the same manner the tones of the basic set produce similar elements. Of course, cadences produced by the distinction between principal and subsidiary harmonies will scarcely be derived from the basic set. But something different and more important is derived from it with a regularity comparable to the regularity and logic of earlier harmony; the association of tones into harmonies and their successions is regulated . . . by the order of these tones. The basic set functions in the manner of a motive. This explains why such a basic set has to be invented anew for every piece. It has to be the first creative thought. (p. 219)

Schoenberg (1948/1975) cites three advantages of composing according to this method. First, because no pitch is repeated more frequently than any other, it prevents any one tone from being interpreted as a tonic and thus references to tonality are avoided.

The construction of a basic set of twelve tones derives from the intention to postpone the repetition of every tone as long as possible. I have stated in my *Harmonielehre* that the emphasis given to a tone by premature repetition is capable of heightening it to the rank of a tonic. But the regular application of a set of twelve tones emphasizes all the other tones in the same manner, thus depriving one single tone of the privilege of supremacy. It seemed in the first stages immensely important to avoid a similarity with tonality. (p. 246)

(For the same reason, he recommends against octave doubling of a tone.)

The second advantage is the unifying effect of employing a basic set and its three mirror transformations (inversion, retrograde, and retrograde inversion). The inversion reverses the pitch direction of each interval in the basic set; the retrograde reverses the temporal order of the tones in the basic set; the retrograde inversion is the reverse temporal order of the inversion. These are called mirror forms because they are symmetrical transformations either in pitch direction (inversion), time (retrograde), or both (retrograde inversion). Schoenberg says about these transformations:

. . . every tone appears always in the neighborhood of two other tones in an unchanging combination which produces an intimate relationship most similar to the relationship of a third and a fifth to its root. It is, of course, a mere relation, but its recurrence can produce psychological effects of a great resemblance to those closer relations. Such features will appear in every motif, in every theme, in every melody and, though rhythm and phrasing might make it distinctly another melody, it will still have the same relationship with all the rest. (Schoenberg, 1948/1975, pp. 246–247)

Elsewhere, he claims that the structure of the tone row is perceived in the mirror transformations just as an object is recognized in different spatial orientations: "[the] mind can operate subconsciously with a row of tones, regardless of their direction, regardless of the way in which a mirror might show the mutual relations, which remain a given quality" (Schoenberg, 1941/1975, p. 223). For him, the use of the basic set and its mirror forms derives from the "law of the unity of musical space."

The third, and final advantage cited by Schoenberg (1948/1975, p. 247) for composing with 12 tones is "that the appearance of dissonances is regulated." He asserts there is no clear distinction between consonant and dissonant intervals.

> What distinguishes dissonances from consonances is not a greater or lesser degree of beauty, but a greater or lesser degree of *comprehensibility*. In my *Harmonielehre* I presented the theory that dissonant tones appear later among the overtones, for which reason the ear is less intimately acquainted with them . . . Closer acquaintance with the more remote consonances—the dissonances, that is—gradually eliminated the difficulty of comprehension . . . The term *emancipation of the dissonance* refers to its comprehensibility, which is considered equivalent to the consonance's comprehensibility. A style based on this premise treats dissonances like consonances and renounces a tonal center. (Schoenberg, 1941/1975, pp. 216–217)

Further, "If dissonances other than the catalogued ones are admitted at all in music, it seemed that the way of referring them all to the order of the basic set is the most logical and controllable procedure toward this end" (Schoenberg, 1948/1975, p. 247).

Despite the apparent restrictions imposed by the use of the tone row, Schoenberg (1941/1975) stresses the tremendous variety afforded by this compositional technique: "One has to follow the basic set; but, nevertheless, one composes as freely as before" (p. 224). And "the possibilities of evolving the formal elements of music—melodies, themes, phrases, motives, figures, and chords—out of the basic set are unlimited" (p. 226). Rufer (1954, p. 79 ff.), a pupil of Schoenberg, provides a summary of factors generating variety within a 12-tone composition. As has already been mentioned, the tone row appears not only in its basic (prime or original) form, but also in its three mirror forms (retrograde, inversion, and retrograde inversion). Moreover, any of the four forms may be transposed to any pitch of the chromatic scale, producing 48 series altogether. The series may be used horizontally (as a melody) or vertically (forming harmonies); the series can be divided between the main part and the accompaniment, or between several contrapuntal parts. Any tone of the series can be placed in any octave, and immediate repetitions of the same tone are permitted. The series can be subdivided into smaller groups (e.g., two groups of six, four groups of three, three groups of four) which can be treated independently (reordered, for example). Other alterations of the series may be required depending on the form, character, and phrasing of a theme. Finally, the pitch materials are shaped and subdivided into phrases through rhythmic means. Thus, despite the adherence to the underlying series, a great number of compositional choices are required.

The Wind Quintet and the String Quartet, No. 4

Traces of 12-tone serialism can found in compositions that were written during Schoenberg's second period of "free atonality," but the Wind Quintet (1924) is one of the first compositions to employ the method in a strict way. The basic series, together with its three mirror forms, provides the materials for all four movements of the piece. Schoenberg (1941/1975, p. 225 ff) uses excerpts from this piece to illustrate various aspects of the style, particularly how the main themes of the piece are derived from the basic set and how the set can be used to generate accompaniment and counterpoint.

The tone row from the Wind Quintet is shown in Figure 1. (Recall that a tone can be sounded in any octave so the octave placement in the notated version is arbitrary.) As Brindle (1966, pp. 9–10) observes, the last tone of the row plus the first 5 comprise one whole-tone scale (E♭ F G A B D♭), and the remaining tones, 6 through 11, comprise another whole-tone scale (B♭ C D E G♭ A♭). A second property of the row is that tones 7 through 12 are related by a perfect fifth to tones 1 through 6. Each tone in the second half is transposed up a fifth from its corresponding tone in the first half, with the exception of the last tone, which is a transposition down a fifth. Stuckenschmidt (1974/1977) notes that, because the two halves relate to each other like tonic to dominant, answers at the fifth above are possible, a property exploited by Schoenberg. "The pseudo-tonal character of these passages makes a paradoxical contrast to the strict 12-tone construction which excludes consonance and tonality" (pp. 295–296). These two properties suggest a natural subdivision of the row into two groups of six tones ("hexachords").

The second piece chosen for the present series of experiments was the String Quartet, No. 4 (1936), a work frequently cited in the music-theoretic literature as demonstrating many features of Schoenberg's 12-tone music. The basic set from this piece is also shown in Figure 1. Only two of its properties will be noted here. The first property is an example of what is known as combinatoriality, whereby an operation (transposition, retrograde, inversion, or a combination of these) on one segment produces a new segment which contains none of the pitches of the original segment. In the present case, if the first hexachord is inverted and transposed down a fifth, the resulting pitches (G A♭ C B E F♯) are all different from those in the first hexachord (and thus identical to the collection of tones in the second hexachord). Schoenberg (1941/1975, p. 225) came to favor rows with this kind

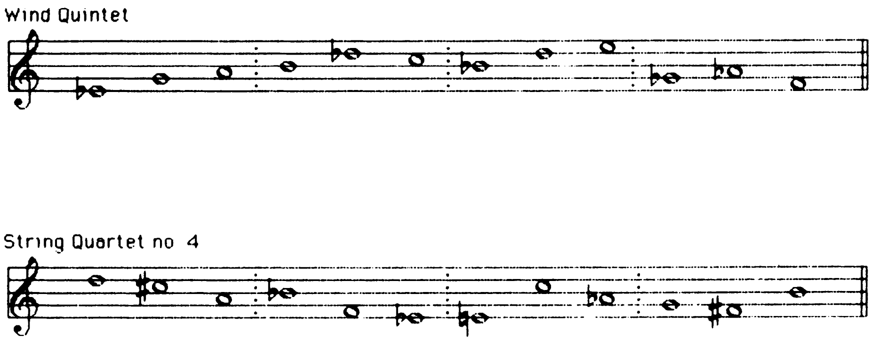

Fig. 1. The tone rows in prime form from the Wind Quintet and the String Quartet, No. 4.

of property because they permitted accompaniment with a transposed inversion without danger of repeating any tone too soon. This property is made use of especially in the String Quartet.

The second property to be noted here of the row from the String Quartet, No. 4 is its local tonal implications. Schoenberg allowed freer use of tonal effects in his serial compositions during the later part of his third compositional period. The rows, in some cases, produce momentary suggestions of tonality. This row can be regarded as suggesting four tonal areas defined by the four groups of three tones ("trichords"). The first trichord suggests A major or perhaps D major or minor; the second suggests B♭ major or E♭ major; the third weakly suggests F minor; and the last is quite ambiguous suggesting B minor, G major, or perhaps E minor. These designations are, of course, approximate because the traditional mechanisms for establishing keys are absent. Schoenberg (1949/1975, pp. 91–92) admits, "In the last few years I have been questioned as to whether certain of my compositions are "pure" twelve-tone or twelve-tone at all. The fact is that I do not know. I am still more a composer than a theorist . . . Whether certain of my compositions fail to be "pure" because of the surprising appearance of consonant harmonies—surprising even to me—I cannot, as I have said, decide."

Overview of the Experiments

Altogether, there were four experiments which were designed with certain objectives in mind. The essential characteristics are summarized in Table 1. Two different tasks were used in the experiments. The first task (used in Experiments 1 and 4) was a probe tone rating task in which a context was followed by a probe tone. Listeners rated how well the probe tone fit with the preceding context in the musical sense of the atonal idiom. The context

TABLE 1
Summary of Experimental Designs

Experiment	Task	Materials
1	Probe tone rating	Neutral: Segments of rows (3, 6, 9, 12 tones)
2	Classification	Neutral: Prime, Inversion, Retrograde, Retrograde Inversion
3	Classification	Excerpts: Prime, Inversion, Retrograde, Retrograde Inversion
4	Probe tone rating	Excerpts: 12-tone Primes

consisted of a complete or incomplete statement of the series from the two pieces in prime form; all 12 tones of the chromatic scale were used as probe tones. This method, introduced by Krumhansl and Shepard (1979), has been used in previous studies with music of other styles and the results of those experiments will be summarized later. In the present study, it was used to assess whether serial contexts impose a hierarchy on the set of chromatic tones and, if so, to determine what factors contribute to the hierarchy. The second task (used in Experiments 2 and 3) was a classification task. The listeners first learned to differentiate between the prime forms of the rows from the two pieces. Then, they were asked to classify all four forms (the prime, inversion, retrograde, and retrograde inversion) according to whether they sound more similar to the prime row of one piece or the prime row of the other piece. The objective of the classification studies was to assess whether listeners perceived the similarity between the prime form of the rows and their mirror forms.

In addition to the two tasks, there were two types of stimulus materials used in the four experiments. In Experiments 1 and 2, the materials were presented in a musically "neutral" form. All tones were presented with the same duration, producing isochronous sequences. In addition, the tones used were "circular" tones with components sounded in five octaves. This produces tones with no clearly defined octave placement. They are called circular because Shepard's (1964) listeners judged tones produced in this way to be related according to the circular dimension of "chroma." In Experiments 3 and 4, the materials were actual excerpts from the two pieces. Thus, they had distinctive rhythms, included immediate tone repetitions, and had unique contours owing to the octave placement of the tones. The rationale behind using the two kinds of materials was to establish certain basic results in the first two experiments which could then be compared to those obtained in the last two experiments using the more musically complex materials.

Several measures were taken to optimize the chance of obtaining consistent and interpretable results. First, the listeners all had extensive musical training, although some had considerably more theoretical knowledge of and experience with atonal music than others. Music backgrounds were assessed using a questionnaire, and considerable care was taken in the analysis of the results to allow for individual differences and relate these to music background characteristics. The same listeners participated in the four experiments, which were always conducted in the same order. This provided listeners with extensive experience with the neutral materials used in the first two experiments before hearing the more complex materials of the last two experiments. Other features of the designs will be noted as the individual experiments are described.

Experiment 1: Probe Tone Ratings with Segments of Tone Rows

The first experiment was a probe tone rating task in which the contexts were segments of the rows from the Wind Quintet and the String Quartet, No. 4 (shown in Figure 1). Previous studies have used this task with a variety of different kinds of contexts. In the first studies (Krumhansl & Shepard, 1979; Krumhansl & Kessler, 1982), the contexts were chosen to strongly indicate a major or minor key; they were diatonic scales, tonic triads, and chord cadences. Following this was a probe tone, which was one of the 12 tones of the chromatic scale. Listeners rated the probe tone as to how well it fit with the preceding context. This process was continued until all 12 tones of the chromatic scale had been presented.

The rating profiles for musically trained listeners in the earlier experiments were consistent with musical intuitions: the tonic received the highest rating, followed by the third and fifth degrees of the scale (which, together with the tonic, form the tonic triad), then the other diatonic scale tones, and finally the nonscale or nondiatonic tones. We have called this pattern the tonal hierarchy and it correlates strongly with the distribution of tones in tonal compositions (Krumhansl, in preparation). Schmuckler (1987) obtained similar results using complex melodic contexts and Palmer and Krumhansl (1987a, b) found influences of tonal hierarchies on judgments of musical phrases.

Two studies have extended the method to music outside the tonal tradition of the common practice period. Krumhansl and Schmuckler (1986b) used as context a passage from Stravinsky's *Petroushka* that employs materials from two different keys at the same time. The purpose of this study was to investigate the capacity to perceive more than one tonal organization simultaneously. Although the probe tone ratings showed contributions of both keys, additional experiments found listeners were unable to focus attention on one component key. Instead, the percept appears to be a complex fusion of the two components.

The other study (Castellano, Bharucha, & Krumhansl, 1984) employed as contexts the themes from ten North Indian rāgs which are based on a variety of underlying scales. One group of listeners had previous training in Indian music while the other group did not. Both groups produced the style-appropriate tonal hierarchies with the highest rating for the tonic (Sa), followed by the fifth scale tone (Pa), the vādi tone (unique for each rāg), then the other scale tones, and finally the nonscale tones. The similarity between the groups was attributed to the fact that the theoretically significant tones were given explicit emphasis in the theme contexts; they were sounded continuously in the drone, and more frequently and with longer durations in the melody.

Although the same method has not previously been applied to atonal music, a related method was used by Temko (1972). That study used recorded excerpts (20 sec in length) from pieces chosen to be representative of the period 1945 to 1970, including a number of serial compositions. Following the excerpt, listeners were required to sing the tone they felt was the most important or prominent pitch in the musical excerpt. The results showed better than chance agreement between listeners on the relatively prominent pitches, suggesting that perceived pitch hierarchies are a general feature of music perception, although the details of the hierarchical ordering will vary from style to style and, indeed, from piece to piece.

In the present experiment, the contexts consisted of the first three, six, or nine tones of the row, or the complete row from the Wind Quintet and the String Quartet, No. 4. These context lengths were chosen to correspond to the segmentation of the rows as used by Schoenberg in the String Quartet; the same context lengths were used for the Wind Quintet to balance the experimental design. The objective of probing the row at different points was to determine whether there were invariant patterns that apply across the entire length of the row and to trace the perceptual organization of the series as it develops.

The principles of 12-tone serialism suggest a number of factors may influence probe tone ratings following these contexts. First, if listeners have internalized the principle that no tone is to be repeated until after all 12 tones of the chromatic scale have been sounded, then ratings for probe tones contained in the incomplete row contexts should be lower than tones not yet sounded. It may be that the more recently a tone has been sounded, the stronger the prohibition and the lower the rating.

A second factor that may be reflected in the probe tone ratings is specific expectations for tones to follow in the series developed through repeated exposure to the rows. The experiment included two replications of the complete design. In the first replication, the contexts were presented in order of increasing length (3 tones, 6 tones, 9 tones, and 12 tones) with all possible probe tones sounded with each context before proceeding to the next context. This provided considerable experience with the order of the tones before the second replication. If learning occurs, then ratings for tones that immediately follow the incomplete contexts should be higher in the second replication than in the first replication.

The final factor concerns tonal effects, and it seems there are two possibilities here. The first is that the compositional technique gives equal emphasis to all 12 tones, that is, there is no hierarchical differentiation between the tones. In this case, the rating profiles might be expected to be flat and not resemble those for any key. An alternative possibility is that listeners understand the technique's intention to avoid a similarity to tonality as the denial of key implications. This might result in low ratings for tones that fit with the local tonal implications of the context, particularly those suggested by the most recently sounded tones. To assess these possibilities, the present results were compared to tonal hierarchies from Krumhansl and Kessler (1982) for major and minor key contexts.

Method

Subjects

Thirteen listeners from the Cornell University community participated in the series of experiments; they were paid at the rate of $4.00 an hour. The experiments were described as investigating the perception of pitch structure in twentieth century music. Each listener had at least five years of formal instruction in instrumental or vocal music and reported being familiar with atonal or 12-tone music. The group included seven individuals who had earned undergraduate degrees in music, four of whom were currently graduate students in the Music Department or had completed graduate degrees in music. The remaining six participants were undergraduate or graduate students in non-music programs at Cornell University. As a whole, the group had an average of 10.5 years of formal instruction in instrumental or vocal music. The primary instruments of the participants were: piano (seven), flute (three), cello and violin (one each), and the remaining listener's training had been in vocal music. They were currently participating in musical activities an average of 8.2 hours per week and listening to music 18.5 hours per week. No listener had absolute pitch.

Apparatus

The stimulus materials were generated on a DMX-1000 signal processing computer (Digital Music Systems) under the control of a PDP-11/23 + computer (Digital Equipment Corporation). The DMX-1000 performs digital synthesis in real time, generates the signal through a 16-bit digital-to-analog converter, and filters the signal at the Nyquist frequency. The signals were amplified by an NAD stereo amplifier (3125), and played at a comfortable listening level through two Mission Electronics Model 70 MK II loudspeakers located on either side of the listener at a distance of approximately 2 feet. Listeners were seated in front of a VT-100 computer terminal (Digital Equipment Corporation) which was used to present written instructions and record responses.

Stimulus Materials

Each trial consisted of a context, a brief silent interval, and then a probe tone. The context consisted of the first 3, 6, 9, or all 12 tones of the row from either Schoenberg's Wind Quintet or his String Quartet, No. 4, as shown in Figure 1. All 12 tones of the chromatic scale were used as probe tones in a randomly determined order. Each tone of the context and the probe tone contained five sinusoidal components at octave intervals, sounded over the five octaves ranging from C_2 (65 Hz) to B_6 (1975 Hz). The amplitudes of the components were determined by a loudness envelope consisting of three parts: a gradually increasing level over the lowest octave and a half, a constant level over the middle two octaves, and a symmetrically decreasing level over the highest octave and a half. This method, patterned after that of Shepard (1964), produces tones that have an organ-like quality with no well-defined highest or lowest pitch; more details of the exact method used can be found in Krumhansl, Bharucha, and Kessler (1982). The tones had linear rise and decay amplitude functions over the first and last 100 msec of their durations. Each tone's duration was 500 msec, as was the silent interval between the context and the probe tone.

Procedure

The following written instructions were given to the participants: "In this experiment you will hear a series of tones drawn from music in the twentieth century atonal idiom. This series will be followed by a single tone. Your task is to rate how well the single tone fits, in the musical sense of the atonal idiom, with the series that preceded it. For this purpose you will use a rating scale from 1 to 7. Use '1' if the last tone fits poorly with the sequence, '7' if the last tone fits well with the sequence, and 2 through 6 for various degrees in between. You should try to develop a criterion for rating that will employ the full range from 1 to 7 on the scale." In addition, the experimenter explained the procedure verbally and demonstrated the task informally at the piano keyboard using both tonal and atonal idioms. In the tonal version, a key was established either harmonically or melodically, and the key-defining contexts were followed by a number of different probe tones. The same was done using a variety of atonal melodies and chord successions. Finally, a short musical example from the atonal idiom, a section from the Gavotte of Schoenberg's Suite for Piano (Op. 25, performed by Paul Jacobs, Nonesuch H-71309), was played to further orient the listeners to the twentieth century idiom.

At the beginning of the first experimental session, listeners heard 24 practice trials which used the row from Krenek's Suite for Violoncello Solo (Op 84); the row for these practice trials consists of the tones: D G♭ F D♭ C B E♭ A B♭ A♭ E G. The context on each trial consisted of the first 3, 6, 9, or all 12 tones of the row, followed by 6 randomly selected probe tones for each context length. Following this, the remainder of the session consisted of eight blocks of 14 trials. The first 2 trials within each block were additional practice trials, and the last 12 trials were experimental trials with each of the 12 possible probe tones represented once. The first four blocks of trials were based on the row from the Wind Quintet with context lengths in the order: 3, 6, 9, and 12 tones. The last four blocks of trials were based on the row from the String Quartet, No. 4 in the same order of context lengths. The second session, usually on the same day, was an exact replication of the eight blocks of trials, except that a different random order of probe tones was used. The two sessions together lasted a total of approximately 1 hour.

Results

Individual Subject Differences

Preliminary inspection of the results revealed a subgroup of listeners producing very similar results to one another. To substantiate this, the responses from each subject for the 192 trials in the entire experiment (excluding the practice trials) were correlated with those for every other subject. Seven listeners all had significant correlations with each other.

That is, all 21 intersubject correlations in this group were significant (at $p < .05$); the average correlation was .35 ($p < .01$). None of these listeners correlated significantly with any of the remaining six listeners. In fact, the average intersubject correlation between the two groups was negative (−.12) and approached significance. Thus, the seven listeners produced results highly consistent with each other and distinctive from the other six listeners. The data for the remaining six listeners were, on average, positively correlated (.06) with one another, although only 3 of the 15 intersubject correlations in this group of six were significant. The intersubject correlations were analyzed using the multidimensional scaling program ALSCAL (Young, Lewyckyj, & Takane, 1983). In both the two- and one-dimensional solutions, there was a compact cluster of seven subjects and a separate, less compact cluster of six subjects.

Accordingly, the listeners were divided into two groups of subjects, called Group 1 (seven subjects) and Group 2 (six subjects). The music background questionnaires were examined for differences between the two groups. In Group 1 were five listeners with university level music degrees, including all four listeners with graduate level training; the two remaining individuals did not have university degrees in music. In Group 2 were two listeners with undergraduate degrees in music and four without. Thus, the two groups tended to differ in terms of the extent of their academic training in music, but this factor did not sharply distinguish between the two groups. They did not differ in terms of the number of years they had been playing and studying instrumental or vocal music or the amount of time they were currently playing or singing. They did tend to differ in terms of the amount of time they were currently listening to music; the average number of hours per week was 22.2 and 14.1 for the two groups, respectively. The most noticeable difference was the amount of time spent listening to "modern classical (atonal)" music in particular; the averages were 3.2 and 0.6 hours per week, respectively. None of these differences, however, approached significance owing to the large variability within groups.

Differences between Replications

In the experiment, there were two complete replications of the entire set of context-probe tone combinations. The first replication provided extensive experience with the tone rows. It is of interest, therefore, to know whether this experience changed the pattern of responding in the second replication. In general, the answer to this question was negative. For Group 1, the correlations between the two replications (with 96 trials each) were .78 and .93 for the Wind Quintet and String Quartet, No. 4 trials, respectively. For Group 2, the corresponding correlations were .61 and .72. These values were highly significant ($p < .01$), indicating that the data from the two replications were similar. Consequently, they will be averaged in the following analyses to increase their stability.

Before averaging, however, an analysis was done to investigate the question of whether listeners learn, through experience, to anticipate the tone that will follow next in the row and give a higher rating to it in the second replication than in the first. Specifically, one can look for a difference between the two replications for the fourth tone of the row when the context contains three tones, the seventh tone when the context contains six tones, and the tenth tone when the context contains nine tones. The ratings for these tones in the first and second replications were entered into an analysis of variance. This analysis showed a tendency for the rating of the next tone to increase (by .55) for Group 1 listeners and to decrease (by .44) for Group 2 listeners. This interaction approached significance [$F(1, 11) = 4.33$, $p = .06$] and cannot be attributed to changes in average overall ratings in the two replications. The effect, however, was small and no other regular differences were found between the two replications, so the following analyses were all based on the average of the two replications.

Serial Position Effects

Figures 2 and 3 show the average ratings given the probe tones by Group 1 and 2 listeners, respectively, for the four contexts lengths and the two pieces. The probe tones are ordered according to their serial position in the row. The tones actually sounded in the context are to the left of the dashed line in each case. As can be seen, there was a strong tendency for Group 1 listeners to give lower ratings to the tones sounded in the context (to the left of the dashed line) than to tones not sounded in the context (to the right of the dashed line). The ratings were especially low for the most recently sounded tones. Group 2 listeners showed the opposite pattern. To substantiate these effects statistically, the entire set of ratings was divided into probe tones sounded in the context and probe tones not sounded in the context. For Group 1 listeners, the average rating for tones in the context was 3.00 and for tones not in the context was 4.61. For Group 2 listeners, the corresponding values were 4.93 and 4.15, respectively. This interaction was highly significant [$F(1, 11) = 38.21, p < .001$].

To look at the effect of how recently a tone had been sounded in the context, the ratings for probe tones contained in the contexts were correlated with the recency value of those tones. That is, if it was the last sounded tone, it would have a recency value of 1; the second to last tone would have a recency value of 2; and so on. The correlation between the ratings and these recency values was, for Group 1 listeners, .64 for the Wind Quintet and .79 for the String Quartet ($p < .01$ for both). Thus, there were consistently lower ratings for more recently sounded tones. The reverse pattern

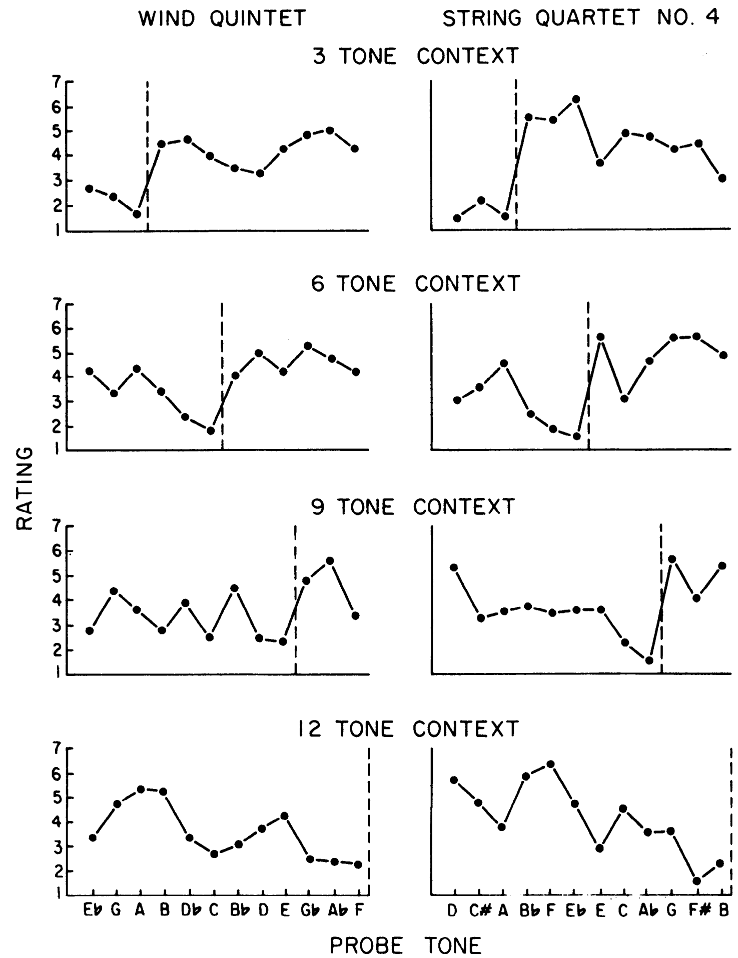

Fig. 2. Probe tone ratings for Group 1 in Experiment 1 for contexts of 3, 6, 9, and 12 tones. The tones on the horizontal axis are ordered according to their position in the tone row. Tones to the left of the dashed lines were contained in the context. Tones to the right of the dashed lines were not contained in the context.

was found for Group 2 listeners; the corresponding correlations were −.75 and −.58 ($p < .01$ for both), indicating that more recently sounded tones received higher ratings.

The next analysis used the ratings for probe tones that were not contained in the contexts. It considered whether there was an effect of the number of positions a tone appears in the series after the end of the context. In other words, did listeners exhibit a tendency to anticipate the next tones in the series? To test this, prospectency values were assigned as follows: 1 if the tone is the next tone following the context; 2 if it is the second tone fol-

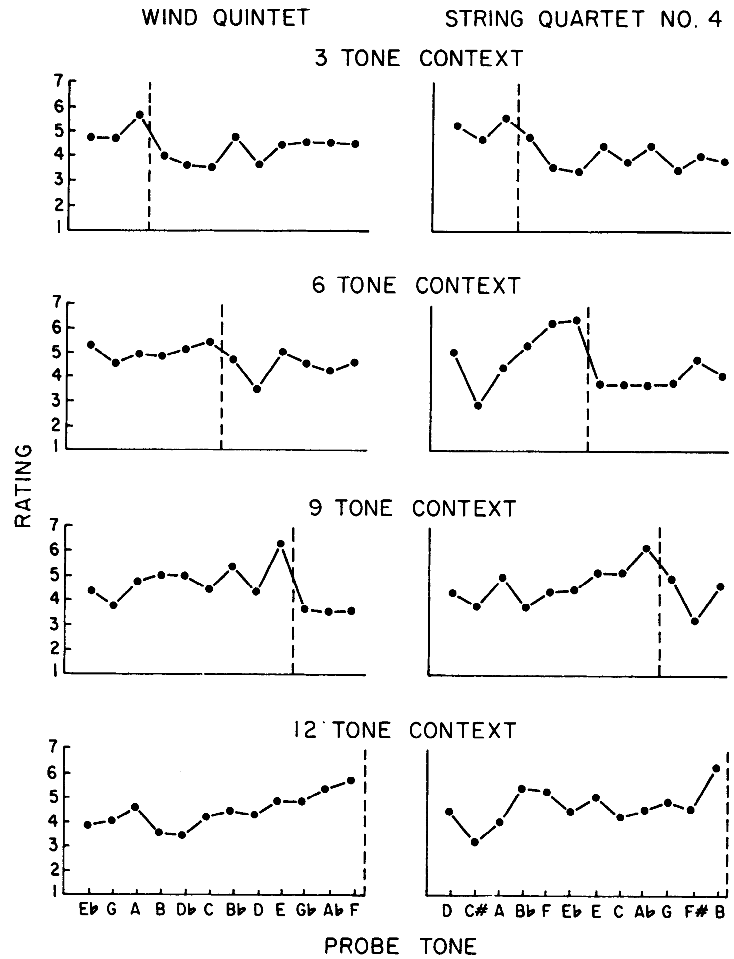

Fig. 3. Probe tone ratings for Group 2 in Experiment 1 for contexts of 3, 6, 9, and 12 tones. The tones on the horizontal axis are ordered according to their position in the tone row. Tones to the left of the dashed lines were contained in the context. Tones to the right of the dashed lines were not contained in the context.

lowing the context; and so on. These values were correlated with the probe tone ratings for tones not contained in the contexts. For Group 1 listeners, the correlations were −.04 and −.42 for the Wind Quintet and String Quartet, respectively. Neither value was significant, but for the String Quartet there was some tendency for tones appearing soon after the context to receive higher ratings. The corresponding correlations for Group 2 were .43 and −.14 for the two pieces, respectively, and neither value was significant. The correlation for the Wind Quintet showed some tendency for tones appearing soon after the context to receive lower ratings.

Tonal Implications

Inspection of Figures 2 and 3 shows considerable variability not accounted for by the order in which tones appear in the rows. The following analysis examined the probe tone ratings for influences of major and minor key tonal hierarchies which may account for some of the residual variability. Because the analysis is fairly complicated, we will refer throughout to two specific examples from the String Quartet, No. 4 in which tonal implications would be expected to be clearer than in the Wind Quintet. Consider first the segment consisting of the first three tones (D C♯ A); these suggest A major (in which they are the first, third, and fourth scale degrees) or the closely related keys of D major or minor (in which they are the first, fifth, and seventh scale degrees). The second example we will consider is the segment consisting of the first six tones (D C♯ A B♭ F E♭). This segment, especially the last three tones, suggests B♭ major (in which the tones are the first, fourth, and fifth scale degrees) or E♭ major (in which the tones are the first, second, and fifth scale degrees).

In an attempt to systematically characterize the tonal implications of the contexts used in the experiment, we used a method similar to the key-finding algorithm reported by Krumhansl and Schmuckler (1986; Krumhansl, in preparation). Basically, the algorithm correlates the distribution of tone durations in a musical sample with the tonal hierarchies of the 24 major and minor keys (using the data from Krumhansl & Kessler, 1982). This gives a quantitative measure of each key's strength (the degree to which the distribution of tone durations in the sample matches the key's tonal hierarchy).

In the present analysis, the tones actually sounded in the context were assigned the value 1 (they were all of equal duration), and the tones not sounded in the context were assigned a value 0. Because nothing about serial position is taken into account, this is called the unweighted model. These values were then correlated with the 24 major and minor key profiles. The obtained correlations gave 24 values, called the key strength vector, for the context in question. For the first three tones of the String Quartet, No. 4, the algorithm found A major to be the strongest key ($r = .67$), followed by D major ($r = .61$) and D minor ($r = .52$). For the first six tones, the strongest key was D minor ($r = .58$), followed closely by B♭ major ($r = .57$). The algorithm could not be applied to the complete 12-tone contexts, because all tones would be assigned values equal to 1, and a correlation could not be computed.

To obtain a key strength vector for all context lengths and to take into account the possibility that more recently sounded tones are psychologically more prominent, an alternative model was also used. It is called the weighted model. In all cases, if a tone was not sounded in the context, it was given the value 0. The first tone in each context was assigned the value 1; the second, 2; and so on. In other words, tones at the end of the context were weighted more heavily. These values were then correlated with the tonal hierarchies of the keys. This model found A major to be clearly the strongest key (with $r = .76$) for the three-tone String Quartet context, and B♭ major to be the strongest key (with $r = .63$) for the six-tone context.

To summarize this first part of the analysis, we have two measures (from the weighted and unweighted models) of the degree to which each major and minor key is suggested by the contexts used in the experiment. In the cases that permit comparisons between the two models, good agreement was found, and the key strength values corresponded quite well with musical intuitions. Finally, as would be expected, the tonal applications for the String Quartet, No. 4 were found to be stronger than those for the Wind Quartet. The average correlation with the strongest key's tonal hierarchy was .57 (unweighted model) and .67 (weighted model) for the String Quartet; the corresponding values for the Wind Quintet were .44 (unweighted model) and .42 (weighted model).

The second part of the analysis considered the degree to which the listeners' rating profiles resembled those for any major or minor key. Each of the rating profiles from this experiment was correlated with the probe tone ratings for all 24 keys (Krumhansl & Kessler, 1982). The 24 correlations were examined for the key whose tonal hierarchy best matched the listeners' ratings. The keys and their corresponding correlations are shown in Table 2 for each of the eight contexts and the two groups of listeners.

In general, Group 1 listeners had some key whose tonal hierarchy correlated significantly with their probe tone ratings. All four contexts from the String Quartet had at least one significant corelation with a key, as did the complete 12-tone context of the Wind Quintet. Two other context lengths (three and nine) for the Wind Quinet narrowly missed the required level of

TABLE 2
Key Whose Tonal Hierarchy Has Highest Correlation with Probe Tone Ratings of Experiment 1

	Group 1		Group 2	
Length	Wind Quintet	String Quartet	Wind Quintet	String Quartet
3	.56 (F♯)	.73 (E♭)*	.30 (g)	.70 (A)*
6	.45 (d)	.63 (e)*	.30 (c)	.68 (B♭)*
9	.56 (D♭)	.71 (b)*	.57 (E)	.46 (a)
12	.67 (G)*	.62 (d)*	.50 (f)	.40 (b)

*$p < .05$.

significance (an r value of .58 is needed). So, the probe tone ratings for Group 1 tended to resemble the tonal hierarchy of some major or minor key. This was less true for Group 2 listeners. Only two of the eight contexts had at least one significant correlation with the tonal hierarchy of some key.

In general, the keys with highest correlations for Group 1 listeners tended to be very distantly related to the keys with the highest correlations for Group 2 listeners, and those for Group 2 listeners corresponded quite well with intuitions concerning tonal implications of the rows. Consider again the three-tone segment of the String Quartet row. Group 1 listeners produced a rating profile that correlated strongly with E♭ major (r = .73), a key very distant from the implied tonal region of A major, whereas this was the key with the highest correlation (r = .70) for Group 2 listeners. Similarly for the six-tone segment, Group 1 listeners produced a rating profile that correlated strongly with E minor (r = .63), a key very distant from the implied tonal region of B♭ major, whereas this was the key with the highest correlation (r = .68) for Group 2 listeners.

The final step of the analysis compared, by correlation, the key strength vectors for the contexts (their tonal implications as quantified by the weighted and unweighted models) and the key strength vectors for the listeners' probe tone ratings (the correlations between their data and the tonal hierarchies). Values will be high to the extent that the probe tone ratings fit with the tonal implications of the contexts. Table 3 shows the values of the correlations. For Group 1 listeners, the correlations were consistently negative and individually significant (at $p < .05$), whether the unweighted or weighted model was used. This means that their probe tone ratings resembled the hierarchy of a key (or keys) that are very distant from the key region suggested by the context. Just the opposite pattern was found for Group 2, although it was somewhat weaker. These listeners tended to produce probe tone ratings consistent with the tonal implications of the contexts, as assessed by either the unweighted or weighted model.

To summarize, there was a very strong pattern that distinguished the two groups of listeners. This pattern is related to tonal implications of the contexts. Group 1 listeners gave low ratings to tones consistent with tonal regions suggested by the context, and high ratings to tones that are inconsistent. This has the consequence that their probe tone ratings resembled keys very distantly related to the key region suggested by the contexts. In contrast, Group 2 listeners produced probe tone ratings that were generally consistent with the tonal implications of the context.

Effects of Context Length

All the analyses so far have been done on the ratings for the eight contexts separately. Inspection of Figures 2 and 3 shows strong effects of the particular context. It seems unlikely, therefore, that there were patterns

TABLE 3
Key Strength Vector for Contexts Correlated with Key Strength Vector for Probe Tone Ratings

	Group 1		Group 2	
Length	**Wind Quintet**	**String Quartet**	**Wind Quintet**	**String Quartet**
		Unweighted Model		
3	−.86*	−.85*	.53*	.96*
6	−.46*	−.73*	.36	.69*
9	−.79*	−.69*	.49*	.26
		Weighted Model		
3	−.87*	−.83*	.55*	.96*
6	−.61*	−.94*	.42*	.85*
9	−.84*	−.84*	.66*	.59*
12	−.58*	−.68*	.79*	.48*

*$p < .05$.

that remained constant across the eight different sets of probe tone ratings. To check this, the probe tone ratings for each context length (3, 6, 9, and 12) were correlated with those for all other lengths. For neither group of listeners was there a single significant positive correlation. What this means is that no consistent patterns appeared independently of context length.

Discussion

The results of this first experiment were characterized by large individual differences in the pattern of responding. Two distinct patterns could be identified. Thus, the first step was to separate the listeners into two groups, called Group 1 and Group 2, exhibiting the two patterns. Examination of the musical background questionnaire showed that Group 1 listeners tended to have more academic training in music and more experience with atonal music, in particular, than Group 2. These factors, however, did not sharply define group membership, so we have refrained from labeling Group 1 the "expert" group, although it contained the listeners more experienced with serial music. Within groups, particularly within Group 1, there was strong intersubject agreement, justifying the analysis of the data for the two groups separately.

The probe tone ratings for Group 1 listeners exhibited a number of characteristics consistent with the idea that they have internalized principles of 12-tone serialism. They gave lower ratings to probe tones that were contained in the incomplete tone row segments, and higher ratings to tones not yet sounded. This is consistent with the requirement that all 12 tones must be sounded before any one is repeated (except by immediate repetition). Ratings for tones sounded most recently were especially low, suggesting these were most salient in perception and, thus, judged as fitting least well with the contexts. These serial position effects had the consequence that no consistent patterns appeared independently of context length.

The probe tone ratings for Group 1 listeners did not, however, reflect specific expectations for tones to follow next in the series. Among tones not included in the incomplete row contexts, there was no consistent serial position effect, that is, no effect of how soon a tone will appear in the series after the incomplete contexts. This would be expected during the first replication, because at no point have they heard the row beyond the point at which it is being probed and because every serial composition employs a unique row. However, this was also true for the second replication when they had had considerable experience in the experiment with the two tone rows. This second replication contained only a very slight increase in the ratings for tones in the series that immediately follow the incomplete contexts as compared with the first replication. This somewhat surprising result suggests limitations in the ability to remember the ordered relation of tones in the series, or alternatively that these listeners did not take this factor into account in forming their judgments.

Finally, the probe tone ratings for Group 1 listeners were such that low ratings were given to tones that are consistent with local tonal implications, and high ratings to tones inconsistent with tonal implications. Rather than producing flat rating profiles, which would be predicted if all tones were perceived as equally prominent, they produced ratings that resembled those for keys very distant from the key region suggested by the contexts. In this sense, then, the pattern is opposite that for tonal contexts. It is impossible to determine from the data whether listeners have adopted a conscious strategy to reverse the tonal hierarchies of traditional music, or whether this style of music engenders a mode of hearing in which tones that deny tonal implications become prominent.

The remaining listeners, comprising Group 2, produced a pattern of results that was in most respects opposite that of Group 1. They gave higher ratings to probe tones contained in the contexts, especially those sounded most recently, and lower ratings to tones not yet sounded. These results are consistent with the fact that probe tone ratings for tonal contexts resemble the distributions of tones in music, and the idea that the most recently sounded tones are the most psychologically salient. Their ratings also showed effects of tonality, with higher ratings for tones consistent with local tonal implications These findings suggest these listeners were processing the musical contexts in terms of knowledge of tonal-harmonic pitch structures. Like the other group of listeners, they showed no evidence of having learned the specific order of tones in the row. Ratings for tones next in the row were no higher in the second replication than in the first; in fact, ratings for these tones showed a slight decrease between the first and second replications. These individual subject differences will be kept in mind as we present the results from the next experiments which used different tasks and materials.

Experiment 2: Classifying Mirror Forms

This experiment used a classification task to investigate whether listeners perceive the similarity between the prime form of the rows and their three mirror forms (inversion, retrograde, and retrograde inversion). A number of studies investigating this question, directly or indirectly, have been reported in the literature. Francès (1972, p. 140 ff) conducted an experiment in which 28 musical examples were written in the 12-tone serial style. Twenty-four examples were based on one series; the four remaining examples were based on another series (which differed from the first series in only the last six tones of the row). The examples included: the prime form of the row, its inversion and retrograde, transposed or not, which were played in a musically neutral way; melodies containing rhythmic variations and octave substitutions; examples using the series harmonically; and, finally, polyphonic examples using two forms of the row simultaneously. Listeners were all professional musicians, but one group had extensive knowledge of serial techniques, having composed in that style.

At the start of Francès' (1972) experiment, listeners were presented with the two series twice. These were followed by the full set of 28 examples. They were required to judge whether each example belonged to the first or the second series. Only for the musically neutral examples was classification accuracy above chance, with somewhat better performance for the more sophisticated listeners. This result, which is most relevant to the present experiment, suggests listeners perceive the retrograde and inversion as similar to the prime forms in the absence of melodic, rhythmic, and harmonic variations. His results for the more complex examples will be considered later in connection with the third experiment where they are more relevant.

Dowling (1972) used a short-term recognition memory task to investigate the perception of mirror forms. Listeners were instructed as to the meaning of the three kinds of transformations and had to judge whether two short melodies were related by the specified kind of transformation. Blocks of trials were run separately for each of the three mirror forms. The melodies consisted of five tones of equal duration, which were randomly generated on each trial subject to certain constraints favoring small intervals. Most listeners had essentially no musical experience. Despite this, performance was above chance for all three mirror forms; the inversion was easiest, the retrograde intermediate, and the retrograde inversion the most difficult. Thus, listeners were able to recognize the mirror forms of these short musical sequences.

DeLannoy (1972), however, conducted a similar experiment in which two sequences were also compared. The sequences were 12-tone series played with equal-duration tones. His listeners, some of whom were music students, were required to judge whether the second sequence was "equal or derived from" the first sequence. Sometimes the second sequence was one of the mirror forms of the first. His listeners did not perform consistently above chance, which might be attributed to the longer sequences in this study than in Dowling's (1972) study or to the vagueness of the instructions to the listeners. He also found in another experiment that listeners were generally unable to indicate the operation a melody had undergone, that is, whether it was an inversion, retrograde, or retrograde inversion. This second experiment used fragments from four 12-tone compositions whose musical complexity might account for his negative results. There are also various anomalies in his treatment of the data, which limit the utility of the study.

Finally, Balch (1981, Experiment 1) considered the psychological effects of the mirror transformations in a somewhat different way. Listeners heard melodies consisting of two parts of equal length. The first part consisted of eight randomly chosen pitches; the second part was an inversion, a retrograde, or a retrograde inversion of the first part, or a different melody (this was the control melody). His listeners, with varying degrees of musical experience, judged how well the second part followed the first. Ratings were higher for the inversion and retrograde than for the control melodies; this difference was not significant for the retrograde inversion. He assumed that these differences in preference judgments reflect the perceived relatedness between the first melody and its inversion and retrograde.

Together these studies suggest listeners may perceive the relation between a pitch sequence and its various mirror forms, especially when they are short or presented in a musically neutral way. There is some suggestion that the ability relates to musical training. In the present experiment, listeners were first trained to assign different labels to the prime form of the rows from the Wind Quintet and the String Quartet, No. 4. Then, all four forms of the rows from the two pieces were intermixed and listeners were asked to assign labels to each sequence according to whether it sounded more like one or the other of the sequences they had learned to label earlier. They were informed as to the nature of the transformations. The materials are shown in Figures 4 and 5; all tones were of equal duration and were "circular" tones (with components sounded in five octaves as in Experiment 1).

Several features of the design favored performance in this task. First, the two rows have very different characteristics. Second, listeners all had exten-

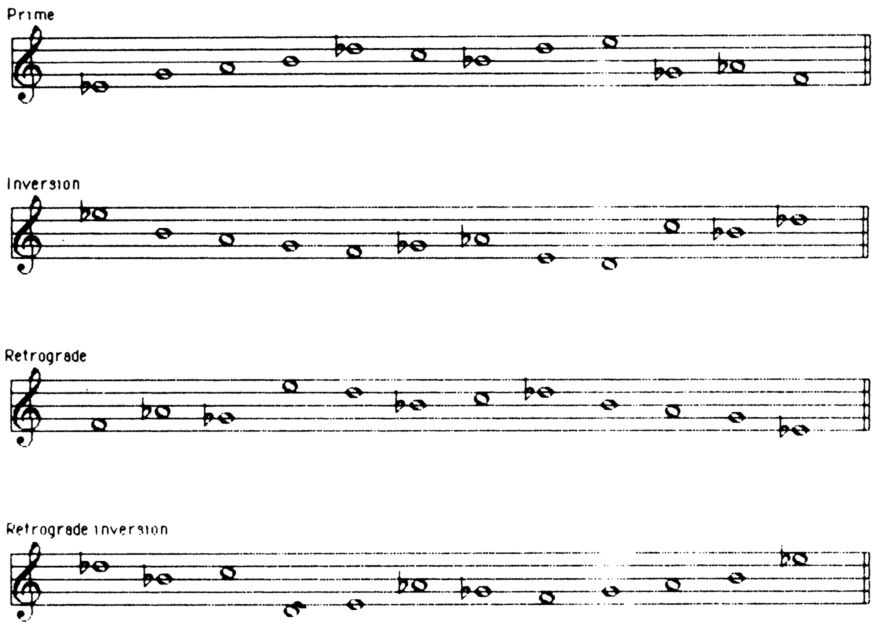

Fig. 4. The prime form of the row from the Wind Quintet and its three mirror forms (inversion, retrograde, and retrograde inversion).

Fig. 5. The prime form of the row from the String Quartet, No. 4 and its three mirror forms (inversion, retrograde, and retrograde inversion).

sive experience with the prime form of the rows in Experiment 1. Third, at the beginning of the present experiment they received as much training as was required to reach a fairly strict criterion of correctly labeling the two prime forms. Fourth, the set of sequences was confined to the two rows and their mirror transformations; sequences based on other rows were not included. Finally, the materials were presented in a musically neutral way; they lacked rhythmic variations or octave transpositions of the tones. These features allow comparisons to be made with the next experiment which used more complex musical excerpts with the same task.

Method

Subjects and Apparatus

Same as Experiment 1.

Stimulus Materials

Each trial consisted of one of the four forms of the rows from the Wind Quintet and the String Quartet, No. 4 shown in Figures 4 and 5. They were presented in the prime form (as in the earlier experiment), the inversion of the prime form (beginning on the same tone), the retrograde of the prime form (without transposition), and the retrograde inversion of the prime form (the untransposed retrograde of the inversion). The tones, which were 500 msec in duration, were produced as in the first experiment. Because the sine-wave components were sounded over five octaves, the octave in which the tones are notated in Figures 4 and 5 is arbitrary.

Procedure

A training session was conducted prior to the experimental trials; it consisted of two parts. In the first, listeners heard the two rows in prime form which were labeled as 1 or 2 for the Wind Quintet and the String Quartet, respectively. They were told that they should try to learn these labels and were permitted to hear the sequences as often as they wished. In the second part of the training session, they were presented with one or the other prime forms, were required to label it as 1 or 2, and were given feedback on their performance. This continued until they reached the criterion that the two prime forms were each correctly identified six consecutive times.

Having completed the training session, listeners were given the following written instructions: "You will be presented with melodic fragments and respond with either 1 or 2. There are now several different fragments rather than just two (as in the training session). Your task is to decide whether the fragment sounds more like number 1 or number 2." In addition, listeners were informed of the nature of the transformations (e.g., inversion, retrograde, and retrograde inversion) they were to hear. After being offered an opportunity to listen to the two prime forms again, listeners proceeded to the experimental trials which included 20 trials of each of the eight sequences shown in Figures 4 and 5 randomly intermixed. The entire experiment lasted approximately 1 hour.

Results

Categorization Accuracy

Figure 6 shows the proportion correct for each of the four forms of the row for the Wind Quintet and String Quartet. As would be expected given the training on the prime forms before the experimental trials, listeners made very few errors categorizing these. The level of categorization accuracy for the other three forms was far above the chance level of performance of .50. An analysis of variance of the proportion correct showed a main effect of form which, however, could be attributed entirely to the difference between the prime form and the other three. When a subsequent analysis was done with the prime forms eliminated there were no overall differences between the three remaining forms. There was, however, a marginally significant interaction between form and subject group [$F(2,22) = 3.18$,

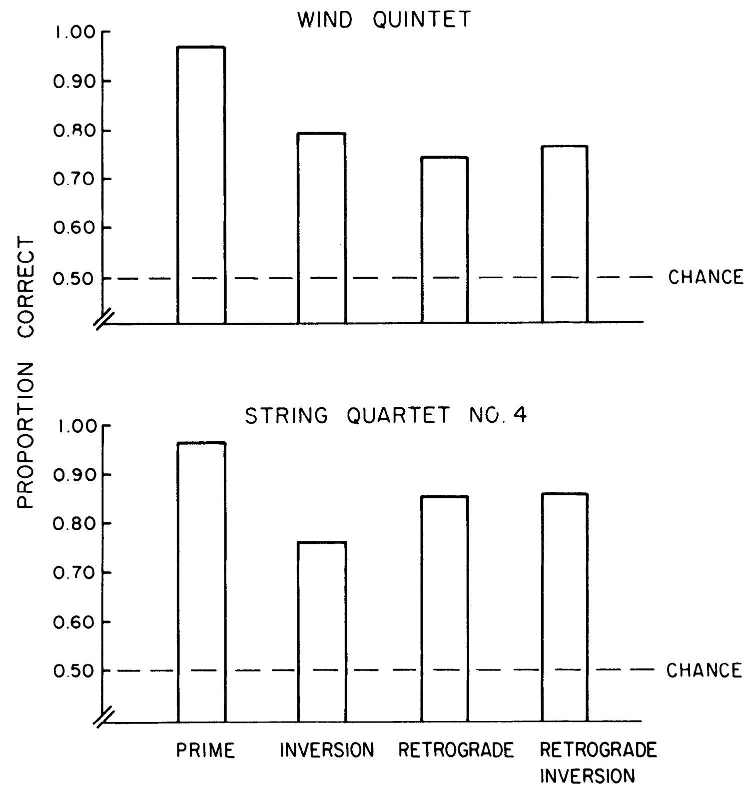

Fig. 6. Proportion correct classification of the prime, inversion, retrograde, and retrograde inversion in Experiment 2.

$p = .06$]; Group 1 listeners more accurately classified the inversion and the retrograde forms than the retrograde inversion, and the opposite pattern was found for Group 2. This interaction is shown in Table 4. No other effects were significant.

Individual Subject Differences

There were fairly large differences across listeners in their level of categorization accuracy. Excluding the prime forms, average performance on the other three forms ranged from .55 to .99. Before considering correlates of these performance differences, it should be emphasized that overall performance was remarkably accurate. All but one of the listeners had a proportion correct that was statistically higher than chance performance (with $p < .05$).

A number of analyses were performed to explore the variation between listeners on performance in this task. The analysis of variance showed no reliable difference in performance between the listeners classified as Group 1 versus Group 2 in Experiment 1. The difference on the nonprime forms between the groups (.82 versus .77) showed somewhat better performance for Group 1 listeners, which is in the direction expected given their more extensive academic music training and their greater familiarity with atonal music. Ranking the listeners in order of performance revealed a regular relationship to the extent of their academic training in music. Of the listeners with the highest eight scores, seven had degrees in music. A variable was constructed as a rough measure of the amount of academic training: 2 = graduate degree in music or candidacy; 1 = undergraduate music degree only; 0 = no music degree. This variable correlated significantly ($r = .73$, $p < .01$) with classification accuracy. There was also some relationship, although not significant, with the number of years playing ($r = .53$) and hours per week listening to atonal music ($r = .46$) and a significant negative relationship ($r = -.61, p < .05$) with hours listening to popular music.

TABLE 4
Proportion Correct Classification in Experiment 2

Form of Row	Group 1	Group 2
Inversion	.83	.73
Retrograde	.85	.74
Retrograde Inversion	.79	.84
Average	.82	.77

Discussion

Listeners in this experiment were able to classify correctly the inversion, retrograde, and retrograde inversion of the rows with their corresponding prime forms. Performance levels were quite high, and all but one listener was significantly more accurate than chance. This was expected given that the design favored performance in various ways noted earlier; the previous literature also generally shows that listeners have this ability for simple sequences. Individual differences were again found, with accuracy correlated with the extent of academic training in music. Listeners classified as Group 1 in the previous experiment (who generally had more academic training) tended to have more difficulty with the retrograde inversion than the other two forms. This is consistent with previous findings and suggests that the difficulty comes from applying two different transformations rather than one. The opposite trend was found for listeners previously classified as Group 2 (with less academic training on average). This may be accounted for because the retrograde inversion is the only mirror transformation that preserves the pitch height direction of the intervals (although their temporal order is reversed).

Experiment 3: Classifying Musical Excerpts

This experiment used the same method as the previous experiment, except the to-be-classified materials were actual excerpts from the Wind Quintet and the String Quartet, No. 4. Thus, they have considerably more melodic and rhythmic variety than the materials used previously. Francès (1972), p. 140 ff) found listeners were unable to classify correctly examples according to their underlying tone row if the melodies contained rhythmic variations and octave transpositions or if the tone row was used harmonically or contrapuntally. This difficulty may result from two features of his experimental design. First, the two underlying rows were very similar, differing only in the last two tones. Second, listeners only had two opportunities to hear the prime form of the rows before the classification task. However, DeLannoy's (1972) listeners also had difficulty hearing the similarity of a musical segment to a complex transformation of it in a task in which two excerpts to be compared were presented in immediate succession and the different underlying rows were chosen to be more dissimilar. These two studies suggest that rhythmic, melodic, and harmonic variations of the row may obscure the perceived relationship between the prime form and its mirror transformations.

A related finding was reported by Pedersen (1975). His materials were based on a 12-tone series which was sounded at the beginning of each trial. After 2 sec a second series was presented that was either identical or different in that two adjacent tones were temporally interchanged. Listeners, who were music students, were required to judge whether the two series were same or different. In one condition, the tones were always sounded in the same octave, and performance was almost perfect. In a second condition, the tones were randomly distributed over a five octave range, and performance was near chance. The sequences in the second condition contained large intervals and diverse contour patterns. It is likely that these factors contributed to the very low performance in the second condition. This finding is difficult to reconcile with the theory of 12-tone serialism that assumes that different octave placements do not change the structure of the row.

The materials in the present study exhibit a wide variety of rhythms, contours, interval sizes and distributions, pitch ranges, and tempi. The excerpts are shown in Figures 7 through 14; there were four excerpts stating each of the four row forms for each of the two compositions. These were randomly intermixed. The experiment evaluates whether listeners can correctly classify these according to their underlying tone row under what would seem like optimal conditions: variants of just these two rows were presented, and listeners were highly familiar with the rows (not only in their prime form but also in their mirror forms) from participating in the previous two experiments.

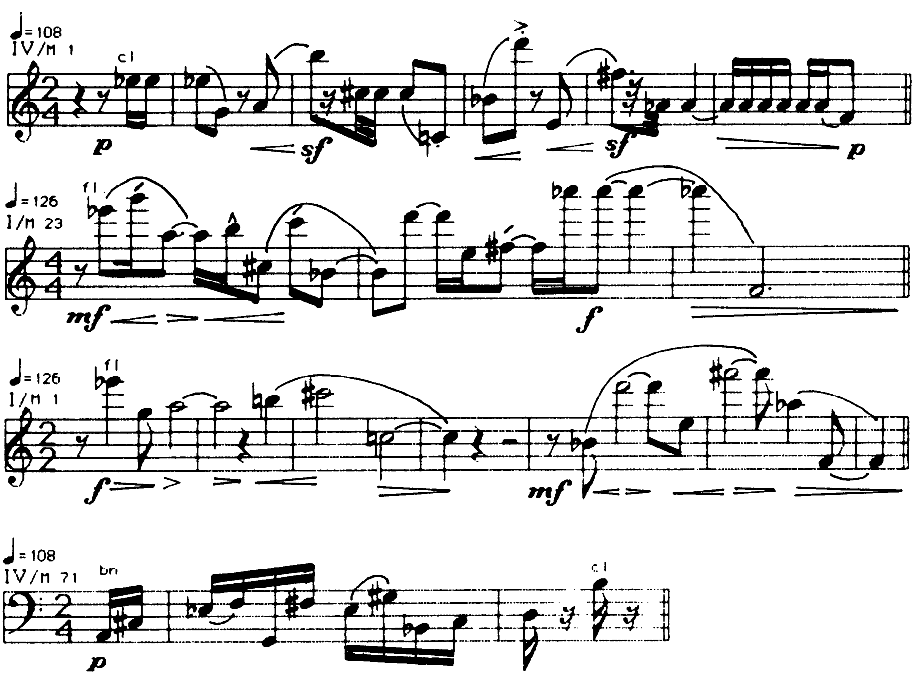

Fig. 7. Excerpts from Wind Quintet with tone row in prime form.

Fig. 8. Excerpts from Wind Quintet with tone row in inversion form.

Fig. 9. Excerpts from Wind Quintet with tone row in retrograde form.

Fig. 10. Excerpts from Wind Quintet with tone row in retrograde inversion form.

Fig. 11. Excerpts from String Quartet, No. 4 with tone row in prime form.

Fig. 12. Excerpts from String Quartet, No. 4 with tone row in inversion form.

Fig. 13. Excerpts from String Quartet, No. 4 with tone row in retrograde form.

Fig. 14. Excerpts from String Quartet, No. 4 with tone row in retrograde inversion form.

Method

Subjects and Apparatus

Same as Experiments 1 and 2.

Stimulus Materials

Thirty-two melodic excerpts were taken from the Wind Quintet and the String Quartet, No. 4; they are shown in Figures 7 to 14. Each consisted of a complete statement of the 12-tone row for that composition in the prime form or one of its three other forms (inversion, retrograde, or retrograde inversion). For each piece, there were four excerpts of each kind used as stimulus materials. Most of the excerpts consisted of uninterrupted passages from a single instrumental part. The four exceptions were cases in which a few (one to three) tones were supplied by another instrument's part in a manner that kept the temporal ordering of the row intact. The excerpts were played at the pitch level indicated in the score (except for the third retrograde from the String Quartet which was shifted down an octave to bring it into the range of the other excerpts). The actual tempi are indicated in the figures. The durations of the excerpts from the Wind Quintet varied from 1.79 sec to 18.23 sec; the durations of the excerpts from the String Quartet varied from 2.91 sec to 8.12 sec.

The tones were produced using a frequency modulation technique (Chowning, 1973). The ratio between the carrier frequency and the modulating frequency was 2:1 for high tones and slightly larger for low tones. The peak of the index of modulation ranged from 4 to 15 depending on the overall amplitude of the tone (determined by dynamics and metrical stress) and the pitch of the tone. During an individual tone, the index of modulation varied linearly with the amplitude envelope. The amplitude envelopes were varied according to the phrasing marked in the music (whether staccato or slurred). This method produced a musically rich timbre, but one that did not clearly sound like any instrumental family.

Procedure

The instructions were the same as in Experiment 2, except that listeners were told that the excerpts they would hear were drawn from the actual pieces of music. Before proceeding to the experimental trials, listeners were given the option to hear the prime forms of the two rows again as in the training trials of Experiment 2. Most listeners took this opportunity. The experimental trials consisted of four blocks in which the 32 melodic segments were randomly ordered. The experiment lasted approximately 45 min.

Results

Categorization Accuracy

Figure 15 shows the average proportion correct for the melodic excerpts of each of the four forms of the rows of the two pieces. For all forms, average performance was well above the chance level of .50. Overall, there was no statistical difference between the four forms of the rows. Unlike Experiment 2, there was no advantage for the prime form. However, again there was a marginally significant interaction between the form and subject group [$F(3,33) = 2.65$, $p = .07$] shown in Table 5; Group 1 listeners had greatest difficulty with the retrograde inversions, whereas Group 2 listeners performed best on these. There were also differences between the individual excerpts. Lower accuracy tended to be found for excerpts of greater contour complexity. The number of changes in contour direction in each excerpt correlated significantly ($r = -.36$, $p < .05$) with classification accuracy. Also, the size (in half steps) of the final interval correlated significantly ($r = -.60$, $p < .01$) with performance. The size of the first interval had no consistent effect, nor did the excerpt's overall duration or average interval size.

Individual Subject Differences

As in the previous experiment, there were fairly large differences across listeners in their categorization performance, with proportion correct ranging from .46 to 1.00. When performance levels in the two experiments were correlated, the highly significant correlation ($r = .92$, $p < .01$) showed that listeners who were accurate in one classification experiment tended to be

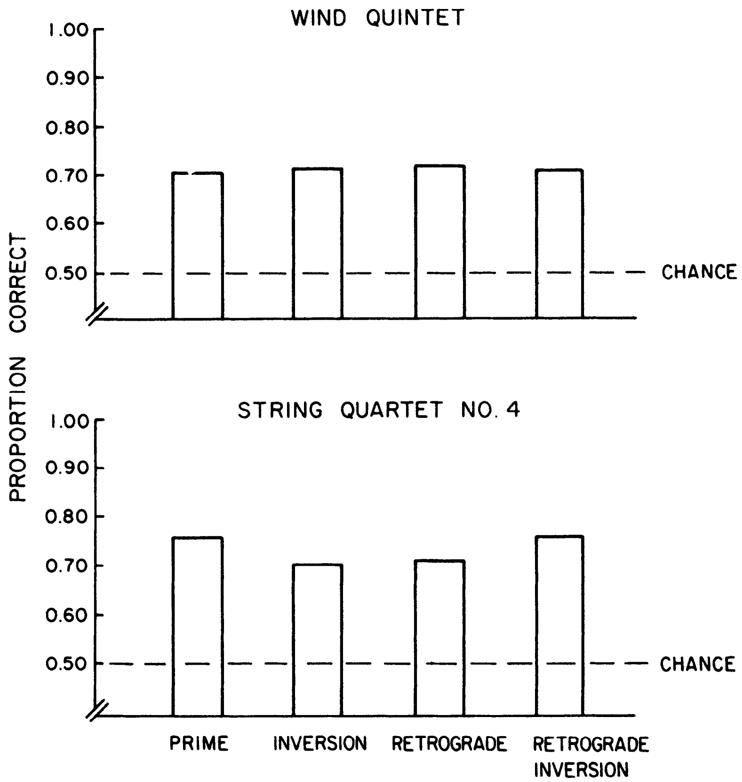

Fig. 15. Proportion correct classification of the prime, inversion, retrograde, and retrograde inversion in Experiment 3.

TABLE 5
Proportion Correct Classification in Experiment 3

Form of Row	Group 1	Group 2
Prime	.82	.64
Inversion	.80	.60
Retrograde	.76	.68
Retrograde Inversion	.75	.73
Average	.78	.66

accurate in the other also. Performance in the present experiment, however, was generally somewhat lower and only 8 of the 13 listeners individually performed reliably above chance.

An analysis of variance of the proportion correct showed no consistent difference between Group 1 and Group 2, as classified in Experiment 1. The average levels of performance were .78 and .66 for the two groups, respectively, and although this difference was large, it was not statistically significant owing to the large variability within groups. Listeners' performance depended on the extent of their academic training in music. The correlation between their performance and a variable coding the amount of their academic training (as in the previous experiment) gave an $r = .68$, $p < .01$. There was also a relationship, although not significant, with the number of hours per week playing ($r = .49$), and a significant negative relationship ($r = -.64$, $p < .05$) with hours listening to popular music.

Discussion

This study found an ability, at least for some listeners, to classify the excerpts according to the underlying row. Performance in this experiment, however, was less accurate than in the previous experiment which used simpler, musically neutral statements of the row and its mirror transforms. There were large individual differences and only somewhat over half of the listeners performed reliably above chance. These individual differences correlated with performance in the previous experiment and the extent of their academic musical training. Listeners previously classified as Group 1 tended to have the greatest difficulty with the retrograde inversion, whereas the opposite was true for the Group 2 listeners. This is the same pattern as in Experiment 2. Prime forms were now classified with approximately the same degree of accuracy as the other forms, suggesting that transfer from the neutral presentation of the prime form (on which listeners were trained) to the prime form excerpts was about as difficult as transfer to the mirror form excerpts.

Some of the individual excerpts were more difficult to classify than others, independently of form. One characteristic consistently related to these differences was contour complexity (the number of changes in pitch height direction), supporting the idea that irregular contours made excepts difficult to relate to the underlying row. The other characteristic was the size of the final interval of the excerpt. When this interval was large, classification performance was low. This suggests one feature listeners used to classify the excerpts was the last interval, and when there were large pitch height separations between the last two tones the interval was difficult to encode. There were no effects of the size of the first interval, average interval size in the excerpt, or total excerpt duration.

Together, these results suggest that the melodic and rhythmic variety of the excerpts made classification with respect to the underlying row difficult, but that some listeners possessed this ability after having extensive experience with the rows in the previous experiments. It is likely that this experience is responsible for the difference in results between this experiment and the experiments reported by Francès (1972) and DeLannoy

(1972), in which listeners were unable to identify musically complex variants of an underlying row.

Experiment 4: Probe Tone Ratings with Musical Excerpts

This final experiment used the probe tone method (Krumhansl & Shepard, 1979) with the prime form excerpts form the Wind Quintet and the String Quartet, No. 4 (Figures 7 and 11). Its main purpose was to evaluate the extent to which the results agreed with those for the musically neutral 12-tone contexts of the first experiment. That is, are there invariant patterns despite the melodic and rhythmic complexity of the excerpts? Are there surface characteristics that influence hierarchies of tone salience in serial contexts? A number of factors were considered, including the serial position of tones in the row, relative and absolute pitch height, relative and absolute tone duration, and number of attacks. We also evaluated the effect of metrical stress by applying the metrical stress grids of Lerdahl and Jackendoff (1983) to the excerpts. Metrical stress values were determined by examining the uninterpreted series of attacks and durations (notated aspects such as grouping of notes by beams and barlines were ignored). The metrical preference rules thus generated a metrical grid which in all cases may not correspond to notated meter. When there were immediate repetitions, the value assigned was the metrical stress for the first attack. And, as in the analysis of the first experiment, we measured the degree to which the ratings conformed to tonal hierarchies of major and minor keys (Krumhansl & Kessler, 1982). Again, individual subject differences were considered.

Method

Subjects and Apparatus

Same as in the previous experiments.

Stimulus Materials

The eight melodic excerpts presenting the rows in prime form from Experiment 3 (Figures 7 and 11) were used as contexts in this experiment. Each melodic segment was followed by a 2-sec pause and a probe tone of 500 msec duration. The probe tones were produced using the same frequency modulation technique as the tones in the context (as described in the Methods section of Experiment 3). They were selected to represent all 12 chromatic scale pitches in an octave range centered (approximately) on the middle of the range of the melodic excerpt.

Procedure

The instructions to the listeners were to rate how well the probe tone fit, in the musical sense of the atonal idiom, with the excerpt that preceded it. They were reminded that this task is the same as that of the first experiment, and told that their judgments should be made similarly. Each of the eight blocks of trials corresponded to one of the eight excerpts. In each block, two practice trials were heard before the 12 experimental trials, with the order of probe tones randomly determined. The experiment lasted about 45 min.

Results

Correspondence with Experiment 1

The first analysis considered the question of whether the probe tone ratings following the melodic excerpts in this experiment were similar to those following the neutral 12-tone contexts of the first experiment. Before examining this question, however, an analysis was done to determine whether any listeners who were previously classified as Group 1 now produced ratings more like those of Group 2 in the previous experiment or vice versa. Each listener's ratings were correlated with the average Group 1 and Group 2 ratings from the first experiment. One listener previously in Group 1 now had a higher correlation with the earlier Group 2 results, and another listener previously in Group 2 now had a higher correlation with the earlier Group 1 results. This shift was confirmed by a multidimensional scaling analysis of the intersubject correlations of the data from the present experiment. However, because there were just two such cases, the listeners in this experiment were not reclassified in the following analyses.

To assess the correspondence between experiments, the groups' average probe tone ratings for each of the eight melodic excerpts were correlated with their ratings from the first experiment for the 12-tone context. For Group 1 listeners, the correlations for all eight excerpts were positive, with an average of .69 ($p < .05$), and all but two of the correlations (Wind Quintet, Excerpt 2; String Quartet No. 4, Excerpt 4) were individually significant. For Group 2 listeners, the correlations between the two experiments were much lower, averaging only .29, and none of the correlations was individually significant. Thus, the responses by Group 1 listeners largely replicated the previous results, whereas Group 2 listeners produced responses that varied considerably from their previous responses.

There was considerable variation in the extent to which individuals' probe tone ratings for the melodic excerpts resembled their ratings from the first experiment. Correlations between the experiments varied from $-.09$ to .49 across individuals. These correlations were compared to listeners' background information. Greater similarity between experiments was found for listeners spending more time listening to atonal music ($r = .68$, $p < .05$) and classical music ($r = .57$, $p < .05$); nonsignificant trends were found of number of years musical instruction ($r = .45$) and the variable coding the extent of their academic training ($r = .40$).

The next analysis looked for a relationship with listeners' performance in the previous two classification experiments. Listeners who produced similar responses in the present and the first experiment were more able to correctly classify the different forms of the row in the second experiment ($r = .54$, $p < .05$) and the melodic excerpts of the third experiment ($r = .72$, $p < .01$). Thus, the same individuals who had performed well in the classification tasks produced probe tone ratings for the melodic excerpts in this experiment that resembled those for the neutral contexts of the first experiment.

Characteristics of the different excerpts were examined to determine whether these affected the degree to which the present results resembled those from the first experiment. No consistent effects were found of the complexity of the contour, the overall duration of the excerpt, or the average interval size within the excerpt for either group of listeners. For Group 1 listeners, this may be because the present results were similar to those of the previous experiment, and thus large differences between excerpts were not found. For Group 2 listeners, there were marked deviations from the previous results, but these appear not to be a regular function of the characteristics of the individual excerpts.

Similarity of Ratings for Melodic Excerpts from the Same Piece

The next analysis considered whether the probe tone ratings for the different excerpts from the same piece were similar despite variations in rhythm and octave placements of the tones. After the ratings were shifted to compensate for different starting pitches in the different melodic excerpts, the rating profiles were correlated. Considerable consistency was found for Group 1 listeners. The average correlation between the rating profiles for the excerpts from the same piece was .60 ($p < .05$) and these were consistently higher than correlations between ratings for excerpts from different pieces as would be expected (these were, in fact, all negative). The same pattern was not found, however, for Group 2 listeners. Their average correlation between ratings for excerpts from the same piece was only .09, and this was no different from the correlations between ratings for excerpts from different pieces. Thus, Group 1 listeners produced ratings that were largely invariant across the different excerpts from a single piece, whereas Group 2 listeners were found to produce highly variable and inconsistent ratings, showing no consistent effect of the piece from which the excerpts were drawn.

Factors Affecting Probe Tone Ratings

Table 6 lists the factors that were considered in trying to determine characteristics of the excerpts that influenced the probe tone ratings for the two

TABLE 6

Correlations between Factors and Probe Tone Ratings in Experiment 4

Factor	Group 1	Group 2
Recency of probe tone in excerpt	.54*	−.18
Absolute pitch height of probe tone in excerpt	.07	.23*
Absolute value of pitch difference between probe tone and average of tones in excerpt	.12	.02
Pitch height of probe tone relative to average of tones in excerpt	−.07	.14
Absolute duration of probe tone in excerpt	−.05	.30*
Duration of probe tone relative to excerpt duration	−.12	.20*
Number of attacks of probe tone in excerpt	−.23*	.12
Metrical stress of first attack of probe tone in excerpt	−.22*	.07

*$p < .05$.

groups of listeners. As in Experiment 1, listeners in Group 1 tended to rate the more recently sounded tones lower. The correlation between their ratings and the recency score (coded with the last tone as 1; the second to last tone as 2; and so on) was .54 ($p < .001$). Of all the factors other than recency, only two had significant correlations for Group 1. They rated probe tones with more attacks (immediate repetitions) in the excerpts lower than tones with fewer attacks ($r = -.23, p < .05$), and tones whose first attack had a stronger metrical stress value lower than tones having a weaker metrical stress ($r = -.22, p < .05$).

For Group 2 listeners, there was a tendency to rate more recently sounded tones higher, as in the first experiment, although the correlation here ($r = -.18$) was not quite significant. The absolute pitch height of the probe tone in the excerpt also correlated positively ($r = .23$, $p < .05$) for Group 2 listeners, which means they gave higher ratings to high tones in the excerpts. They also gave higher ratings to tones sounded for longer absolute durations ($r = .30$, $p < .01$) and longer durations relative to the excerpt's total duration ($r = .20, p < .05$). No other factors had a consistent relationship to Group 2 listeners' responses. Although the majority of correlations in this table were not significant, it is interesting to note that the signs of the correlations were opposite for the two groups of listeners in most cases.

Tonal Implications

The final analysis examined the probe tone ratings for influences of major and minor key tonal hierarchies. Each of the rating profiles was shifted to compensate for the different starting pitches; they will be treated as though the starting pitches are those shown in Figure 1 (E♭ for Wind Quintet; D for String Quartet, No. 4). The resulting rating profiles were then correlated with the probe tone ratings for all 24 major and minor key contexts (using the data from Krumhansl & Kessler, 1982). The 24 correlations were examined for the key whose tonal hierarchy best matched the listeners' ratings. The keys and their corresponding correlations are shown in Table 7 for the eight excerpts and the two groups of listeners.

For Group 1, seven of the eight excerpts produced results correlating significantly with the profile of some key. Moreover, examining the table shows considerable consistency in the keys for the different excerpts from the same piece. For the Wind Quintet, the keys with the highest correlations were E minor and B minor (which is the next key on the circle of fifths from E minor). Similarly for the String Quartet, No. 4, the keys with highest correlations were B♭ major, F major (the next key around the circle of fifths from B♭ major) and B♭ minor (the parallel minor of B♭ major). Thus, for this group, the probe tone ratings showed consistent similarities to profiles for major and minor keys. In this measure as in other measures, Group 2 listeners showed considerable variability. Only one of the eight excerpts produced results that correlated significantly with the profile for any key and there was little consistency across the excerpts from the same piece.

As in the analysis of the first experiment, the excerpts were evaluated in terms of their tonal implications. Because these eight excerpts contained all 12 tones, it was necessary to use the weighted model (with more recent tones weighted more heavily). The resulting 24 dimensional vector of key strengths was compared with the 24 correlations between the listeners' ratings and the major and minor key profiles. As in Experiment 1 (Table 3), Group 1 listeners showed consistently negative correlations. They averaged $-.67$ and each was individually significant (at $p < .01$). Thus, tones fitting with the tonal implications of the contexts (with the final tones weighted more heavily) were rated lower than notes not fitting with the tonal impli-

TABLE 7
Key Whose Tonal Hierarchy Has Highest Correlation with Probe Tone Ratings of Experiment 4

	Group 1		Group 2	
Excerpt	Wind Quintet	String Quartet	Wind Quintet	String Quartet
1	.71 (b)*	.81 (B♭)*	.49 (g♯)	.49 (E)
2	.58 (e)*	.67 (F)*	.49 (E)	.41 (c)
3	.63 (e)*	.71 (b♭)*	.42 (f♯)	.33 (F)
4	.79 (e)*	.39 (b♭)	.66 (G)*	.47 (F)

*$p < .05$.

cations. Group 2 listeners again produced inconsistent results; for two of the eight excerpts there was a significant correlation between the key strength vector and the correlations of the rating data with the major and minor keys; one of the significant correlations was negative and the other was positive. The values averaged .09.

Discussion

The main finding of this last experiment was considerable consistency for Group 1 listeners between the probe tone ratings for the musical excerpts and the probe tone ratings for the neutral tone row contexts of the first experiment. Thus, the perceived hierarchy of tone salience was largely unaffected by the rhythmic and melodic variations of the musical excerpts. As before, they gave lower ratings to tones sounded more recently in the contexts than tones sounded earlier, and lower ratings to tones that fit with the tonal implications of the context than tones that deny these implications. These listeners also showed considerable consistency between the four different excerpts of each of the two pieces. Thus, there is strong evidence for invariant perceptual effects of the structure of the row despite the melodic and rhythmic variations of the excerpts.

Of the remaining factors considered, only two had consistent effects for Group 1 listeners. First, lower ratings were given to tones with more attacks (immediate repetitions) than to tones with fewer attacks. Second, they gave lower ratings to tones that appeared at metrically stressed points in time than to tones that appeared at unstressed points. Both these results were opposite what would be expected for tonal contexts, in which immediate repetitions and metrical stress would generally contribute to tone salience. Thus, to the extent that surface features influenced probe tone judgments of listeners in this group, the effects were opposite those expected if the materials were being perceived in a tonal manner.

The results for Group 2 listeners were characterized by a great deal of variability. Their ratings in this experiment did not correlate with those in the first experiment, although in both they tended to rate tones sounded more recently in the context higher than tones sounded earlier. Nor did these listeners produce consistent patterns across the four different excerpts of each of the two pieces. To the extent that it was possible to identify factors influencing their judgments, they appeared to respond to very obvious features of the excerpts. They gave generally higher ratings to high tones than to low tones, and to tones with longer (absolute or relative) durations than to tones with shorter durations. Effects of tonal implications were inconsistent, as were effects of metrical stress patterns or any other factor considered. It appears, then, that the melodic and rhythmic variations of the excerpts had a large effect on the salience of the tones that masked the similarities between the different excerpts and between the excerpts and the neutral 12-tone rows of the first experiment.

Summary and Conclusions

We view this series of experiments as an exploratory study of the perception of 12-tone serial music. The general question addressed was the nature of perceptual representations of pitch structures in this musical style. More specifically, have listeners internalized the principle that serial composition requires the exhaustive employment of the entire collection of chromatic tones? Do they learn to anticipate successive tones in the row? Are they able to hear the music without forming associations to tonal structures of the music in which they have been deeply acculturated? Are the mirror transformations perceived as similar to the prime form of the row? Are there perceptual invariances despite rhythmic and melodic transformations of the row in the actual music?

The answer to these questions is a highly qualified "yes." The primary qualification has to do with large individual differences that were found throughout the entire series of experiments. The results for only some of the listeners were those that might be predicted from 12-tone theory. The second set of qualifications come from the fact that the experiments were biased in favor in favor of obtaining positive results. Very simple materials and tasks were employed, and listeners were provided by the experiments with extensive experience with a restricted set of sequences. Finally, not all of the effects expected based on 12-tone theory were obtained for any of the listeners, even listeners who showed other effects consistent with the theory.

One group of listeners could be identified who produced a cluster of results generally consistent with theoretical predictions. Group membership correlated with more academic music training and experience with atonal music, but neither these background characteristics nor any others sharply differentiated between these listeners and the others. This final section will focus on the results for this first group of listeners because of their greater theoretical interest, and will emphasize general patterns that emerge from the four experiments.

One of the most consistent findings of the first probe tone study for these listeners was the effect of the tone's serial position in the rows. Tones sounded in the incomplete statements of the rows were rated as fitting less well with the context than tones not yet sounded, consistent with the principle requiring constant circulation of the 12 chromatic scale tones. This effect had the consequence that no patterns were invariant across the four context lengths (3, 6, 9, or 12 tones). Moreover, among the tones contained in the contexts, lowest ratings were given to those sounded most recently, as though the prohibition against their repetition is strongest. This effect of tone recency was also very reliable in the second probe tone study which used musical excerpts from the two pieces as contexts, and again seems to follow from the requirement to sample the entire chromatic collection.

These listeners' probe tone ratings, however, did not reflect specific expectations for tones that follow next in the series after the incomplete statements of the rows. This was true despite opportunities provided by the experiment to become familiar with the two rows. It is possible, of course, that these listeners simply chose not to let this factor influence their ratings. The plausibility of this is weakened by the finding that tones not yet sounded in the context had generally higher ratings than tones already sounded. This indicates that listeners do take serial order into account in a very general way in forming their judgments. At least, they reliably differentiated between tones at the beginning of the row (and contained in the incomplete contexts) and tones not yet sounded.

The alternative possibility is that these listeners did not learn the precise ordering of tones in the row. This possibility is consistent with results showing that sequences not conforming to tonal structure are difficult to remember. This has been found for both melodies (e.g., Cuddy, Cohen, & Mewhort, 1981; Cuddy, Cohen, & Miller, 1979; Francès, 1972, pp. 79 ff) and chord sequences (Bharucha & Krumhansl, 1983; Krumhansl & Castellano, 1983). These studies show tonal structure confers considerable advantages to pitch memory. On the other hand, Pedersen's (1975) listeners were able to detect temporal reorderings of tones in 12-tones series sounded in a single octave. These changes, however, were accompanied by contour changes, a factor known to be perceptually salient (e.g., Dowling, 1978) and it may be contour, rather than memory of precise temporal order, that is responsible for Pedersen's result. Thus, the possibility that these listeners did not develop precise memory representations for tone order in the series must be taken seriously.

The next finding concerned tonal implications as they were reflected in the probe tone ratings. The first group of listeners in both probe tone studies produced a very interesting pattern: lower ratings for tones fitting with local key suggestions than for tones that deny these suggestions. On first consideration, this result could be taken to mean that listeners have succeeded in freeing the tones from tonal associations, permitting a new organization to emerge. It is also in general agreement with the finding of Castellano et al. (1984) that listeners familiar with Western music can set aside their knowledge of major and minor key hierarchies and perceive the hierarchies appropriate to a novel musical style. On second consideration, however, the result suggests a strategy, whether conscious or not, of relating the serial contexts to tonal hierarchies of major and minor keys, and simply reversing the ordering of the ratings. In fact, a few listeners commented on having used this kind of response rule. If this was the strategy used, it suggests that diatonic tonal hierarchies were still activated, but that these were used in a way opposite to tonal music.

In connection with these effects of tonal hierarchies, the question also arises whether they are independent of the serial position effects summarized earlier. The answer to this is not clear. The presence or absence of tones in the contexts enters into the models used to characterize key implications of the contexts, and tone recency enters into the weighted model in particular. The similar results of the weighted model to the unweighted model (in which recency is not taken into account), suggests that the recency effect is not critical to the conclusions concerning tonal effects. Presence versus absence of tones in the contexts, however, cannot be eliminated as a contributing factor. The general correspondence between tone distributions and tonal hierarchies (Krumhansl, in preparation; Castellano et al., 1984) suggests that, except under highly artificial circumstances, it will be impossible to separate these two factors completely. Thus, the effects in this study attributed to tonal implications would seem properly considered as such, rather than derivative solely of the serial position effects also found in these studies.

The last effect found in the probe tone studies for this first group of listeners was the notable similarity between the different excerpts from the same composition, and between these and the neutral 12-tone contexts. This similarity indicates that the underlying organization of the sequences was perceived as largely unaffected by the rhythmic and melodic variations of the excerpts. In particular, the probe tone ratings were unaffected by the tone's absolute or relative pitch height, or its absolute or relative duration in the excerpts. To the extent that surface characteristics influenced the ratings, the effects were opposite those that would be expected for tonal music: lower ratings were given to tones with more attacks and which appeared at metrically stressed points in the excerpts. Thus, just as listeners in this group reversed the normal tonal hierarchies, they also reversed the effect of these characteristics from what would be expected for tonal music.

Finally, this group performed relatively well in the classification experiments which examined the ability to classify the mirror transformations (inversion, retrograde, and retrograde inversion) with the corresponding prime forms. Above chance accuracy was, in fact, found for most listeners when the materials were presented in a musically neutral manner. However, the listeners in this first group performed almost as well with the excerpts from the two pieces as with the neutral sequences. This finding reinforces the conclusion from the probe tone studies that these listeners perceived the structure of the row as invariant despite a wide variety of transformations including the reversal of temporal order, pitch direction, and the application of both these operations.

The results for the second group of listeners were, in most respects, opposite those just described for the first group of listeners. A number of findings suggested they tended to interpret the pitch sequences according to traditional tonal principles. Probe tone ratings favored tones sounded in the contexts, especially those sounded most recently. They also showed patterns conforming to local tonal implications. Various results indicated these listeners had difficulty responding to the materials in a consistent way. The probe tone ratings for the musical excerpts bore little resemblance to those for the neutral statements of the rows, and were influenced by such obvious surface features of the excerpts as pitch height and tone duration. These listeners were able to classify the mirror transformations with the appropriate prime form only with the neutral materials; accuracy dropped markedly when musical excerpts were used. The empirical literature (e.g., Deutsch, 1972) suggests that octave displacements may have contributed to these inconsistencies; changes in rhythm and phrasing may also have had an effect which is supported by the influence found of tone duration.

To return to the theoretical issues with which we began, the experiments were most successful in demonstrating the unifying potential of basing the materials of an entire composition on a single tone row. Rhythmic, melodic, and mirror transformations did not, for some listeners, mask the common origin of the variants in the underlying row. The principle of serial music requiring the constant circulation of all 12 chromatic scale tones also received empirical support—but, again, only for some listeners. The support for this principle does, however, suggest that serial contexts may engender expectations for tones that will fill out the chromatic collection. As to whether listeners can suppress associations to tonal pitch structures, the results were less conclusive. Certainly, the data for some listeners did not conform to patterns found in similar studies with tonal music. But, the fact that the results appeared to be simple reversals of tonal effects suggests knowledge of tonal structures was still operative rather than an entirely different perceptual organization.

This study is obviously limited in a number of ways. The materials were derived from two compositions by Schoenberg and are by no means representative of the vast diversity of compositions written in the style of 12-tone serialism. Second, the materials and perceptual tasks were relatively simple and do not address many issues that arise from this style; harmony in serial music is a question of special psychological interest. Finally, we have confined our attention to theoretical statements by Schoenberg about the compositional technique as they relate to the experimental findings. We have not considered more formal structural descriptions an introduction to which can be found in Forte (1973), Perle (1981), Rahn (1980), and Wuorinen (1979, which is of special interest for his treatment of serial techniques in perceptual terms). Despite these limitations, the present study has demonstrated the applicability of psychological methodologies to serial music and found patterns of results some of which confirm perceptual correlates of principles of serial composition.[1]

References

Balch, W. R. The role of symmetry in the good continuation ratings of two-part tonal melodies. *Perception & Psychophysics,* 1981, ***29,*** 47–55.

Bharucha, J. J., & Krumhansl, C. L. The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition,* 1983, ***13,*** 63–102.

Brindle, R. S. *Serial composition.* London: Oxford University Press, 1966.

Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. Tonal hierarchies in the music of North India. *Journal of Experimental Psychology: General,* 1984, ***113,*** 394–412.

Chowning, J. M. The synthesis of complex audio spectra by means of frequency modulation. *Journal of the Audio Engineering Society,* 1973, ***21,*** 536–544.

Cuddy, L. L., Cohen, A. J., & Mewhort, D. J. K. Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance,* 1981, ***7,*** 869–883.

Cuddy, L. L., Cohen, A. J., & Miller, J. Melody recognition: The experimental application of musical rules. *Canadian Journal of Psychology,* 1979, ***33,*** 148–157.

DeLannoy, C. Detection and discrimination of dodecaphonic series. *Interface,* 1972, ***1,*** 13–27.

Deutsch, D. Octave generalization and tune recognition. *Perception & Psychophysics,* 1972, ***11,*** 411–412.

Dowling, W. J. Recognition of melodic transformations: Inversion, retrograde, and retrograde inversion. *Perception & Psychophysics,* 1972, ***12,*** 417–421.

Dowling, W. J. Scale and contour: Two components of a theory of memory for melodies. *Psychological Review,* 1978, ***85,*** 341–354.

Forte, A. *The structure of atonal music.* New Haven: Yale University Press, 1973.

Francès, R. *La perception de la musique* (2nd edition). Paris: J. Vrin, 1972.

Krumhansl, C. L. (in preparation) *Tonal structures and music cognition.* New York: Oxford University Press.

Krumhansl, C. L., Bharucha, J. J., & Kessler, E. J. Perceived harmonic structure of chords in three related musical keys. *Journal of Experimental Psychology: Human Perception and Performance,* 1982, ***8,*** 24–36.

Krumhansl, C. L., & Castellano, M. A. Dynamic processing in music perception. *Memory & Cognition,* 1983, ***11,*** 325–334.

Krumhansl, C. L., & Kessler, E. J. Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review,* 1982, ***89,*** 334–368.

Krumhansl, C. L., & Schmuckler, M. A. Key-finding in music: An algorithm based on pattern matching to tonal hierarchies. Paper presented at the Mathematical Psychology Meeting, 1986a.

Krumhansl, C. L., & Schmuckler, M. A. The *Petroushka* chord: A perceptual investigation. *Music Perception,* 1986b, ***4,*** 153–184.

1. This research was supported by Grant MH 39079 from the National Institute of Mental Health to the first author. We are grateful to Mark LaPorta for advice on musical matters, Jim Grant for preparation of the musical figures, and Fred Lerdahl for comments on an earlier draft.

Krumhansl, C. L., & Shepard, R. N. Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance,* 1979, *5*, 579–594.

Lerdahl, F., & Jackendoff, R. *A generative theory of tonal music.* Cambridge: M.I.T. Press, 1983.

Palmer, C., & Krumhansl, C. L. Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance,* 1987a, *13*, 116–126.

Palmer, C., & Krumhansl, C. L. Pitch and temporal contributions to musical phrase perception: Effects of harmony, performance timing, and familiarity. *Perception & Psychophysics,* 1987b, *41*, 505–518.

Pedersen, P. The perception of octave equivalence in twelve-tone rows. *Psychology of Music,* 1975, *3*, 3–8.

Perle, G. *Serial composition and atonality* (5th edition). Berkeley: University of California Press, 1981.

Rahn, J. *Basic atonal theory.* New York: Longman, 1980.

Rufer, J. *Composition with twelve tones.* London: Rockliff, 1954.

Schmuckler, M. A. Expectations in music: Additivity of melodic and harmonic processes. Unpublished Ph. D. dissertation, Cornell University, 1987.

Schoenberg, A. *Theory of harmony.* (*Harmonielehre,* R. E. Carter, trans.). Berkeley: University of California Press, 1922/1978.

Schoenberg, A. Composition with twelve tones (1). In L. Stein (Ed.), *Style and idea.* London: Faber and Faber, 1941/1975.

Schoenberg, A. Composition with twelve tones (2). In L. Stein (Ed.), *Style and idea.* London: Faber and Faber, 1948/1975.

Schoenberg, A. My evolution. In L. Stein (Ed.), *Style and idea.* London: Faber and Faber, 1949/1975.

Shepard, R. N. Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America,* 1964, *36*, 2346–2353.

Stuckenschmidt, H. H. *Arnold Schoenberg: His Life, World and Work.* (*Schoenberg, Leben, Umwelt, Wert,* H. Searle, trans.). New York: Schirmer, 1974/1977.

Temko, P. M. The perception of pitch predominance in selected musical examples of avant-garde composers: 1945–1961. Unpublished Ph. D. dissertation, Florida State University, 1971. *Dissertation Abstracts,* 1972, *32*(9), 5275A–5276A.

Wuorinen, C. *Simple composition.* New York: Longman, 1979.

Young, F. W., Lewyckyj, R., & Takane, Y. The ALSCAL procedure. In *SUGI Supplemental Library User's Guide.* Cary, NC: SAS Institute Inc., 1983.