Independent Temporal and Pitch Structures in Determination of Musical Phrases

Caroline Palmer and Carol L. Krumhansl
Cornell University

In two experiments we addressed the roles of temporal and pitch structures in judgments of melodic phrases. Musical excerpts were rated on how good or complete a phrase they made. In Experiment 1, trials in the temporal condition retained the original temporal pattern but were equitonal; trials in the pitch condition retained the original pitch pattern but were equitemporal; and trials in the melody condition contained both temporal and pitch patterns. In Experiment 2, one pattern (pitch or temporal) was shifted in phase and recombined with the other pattern to create the pitch and temporal conditions. In the melody condition, both patterns were shifted together. In both experiments, ratings in the temporal and pitch conditions were uncorrelated, and the melody condition ratings were accurately predicted by a linear combination of the pitch and temporal condition ratings. These results were consistent across musicians with varying levels of experience.

In music perception, highly complex auditory information is analyzed and interpreted by the listener. The sounded events are heard as organized into well-formed, coherent structures such as musical phrases. Each event can be characterized by its pitch, duration, loudness, and timbre. Perceptually, these variables give rise to larger, more abstract patterns such as tonality and rhythm. The experiments reported here focus on two of these attributes: pitch and temporal structure. These can be defined independently of one another in the music; the temporal structure can be described without reference to pitch and vice versa. But these attributes do not necessarily have independent influences on the perception of musical phrases. It may be, for example, that the listener's organized percept results from an interaction of these, and possibly other, variables and cannot be accounted for solely in terms of the separate attributes or their additive combination. More specifically, in this study we investigate the question of whether musical phrase structure can be described in terms of tonal and rhythmic hierarchies arising from the pitch and temporal information.

The notion of hierarchical organization is at the core of most models of tonal and rhythmic organization (Cooper & Meyer, 1960; Deutsch & Feroe, 1981; Krumhansl & Kessler, 1982; Lerdahl & Jackendoff, 1983). Hierarchical models describe the way individual elements combine into larger structural units. As a description of psychological processes, hierarchical organization provides an effective way of encoding, interpreting, and

predicting musical events. According to empirical and theoretical accounts of pitch structure (Krumhansl & Kessler, 1982; Krumhansl & Shepard, 1979; Lerdahl & Jackendoff, 1983), listeners perceive pitches hierarchically, with certain pitches subordinate in salience or importance to others. Pitches that dominate in the hierarchy are perceived as more closely related to one another than to pitches lower in the hierarchy (Krumhansl, 1979). In addition, the contextually important tones are better remembered and infrequently confused with less important tones (Bharucha, 1984; Krumhansl, 1979).

Many models of rhythmic organization (Cooper & Meyer, 1960; Jones, 1976; Lerdahl & Jackendoff, 1983) are also hierarchical in nature, with points of relatively strong and weak stress. In addition, each of these models assumes a nested structure in which smaller temporal segments combine into larger phrases. In Western tonal music, patterns of equal time intervals and regular accent placements are found, sequences of which are called metrical. In empirical studies, both musically trained and untrained subjects more accurately reproduce metrical sequences than nonmetrical sequences (Essens & Povel, 1985; Povel, 1981). Similarly, Deutsch (1980) found increased recall for melodies that have regularly recurring pitch and temporal patterns.

The question of how pitch and temporal structures work in combination has received less theoretical and empirical analysis. Theoretical accounts often treat the two components separately. For example, models of rhythm (Longuet-Higgins & Lee, 1982; Martin, 1972; Povel, 1981) have been proposed without regard for the influence of pitch information. Similarly, models of pitch structure (Deutsch & Feroe, 1981; Krumhansl, 1979; Shepard, 1982) do not take rhythmic influences into account. The proposal of Simon and Sumner (1968) also treats the two components separately. Other, primarily music-theoretic, descriptions (Benjamin, 1984; Cooper & Meyer, 1960; Lerdahl & Jackendoff, 1983; Tenney & Polansky, 1980) assume that rhythmic organization is at least partially influenced by pitch information. These proposals, which have not been subjected to em-

pirical analysis, do not give explicit rules for weighting the relative influence of pitch and temporal factors, nor do they specify whether the factors combine interactively or independently.

Empirical studies typically emphasize either temporal or pitch structure without considering the effect that one factor might have on the other. Experimental results are often based on stimulus materials that are equitemporal (pitches of equal duration) or equitonal (temporal patterns formed of uniform pitches). There are a few studies, however, that have considered both factors together. Recently, Monahan and Carterette (1985) obtained similarity ratings for pairs of melodies. They found that the major dimensions governing perceived melodic similarity reflected rhythmic features of the melodies; pitch-based features had weaker effects. They also obtained results showing a negative trade-off between the weights given to rhythmic and pitch dimensions. Listeners emphasizing one kind of information tended to deemphasize the other. These results suggest that rhythmic and pitch structures are perceived separately and that attention is allocated to one aspect of the stimulus at the expense of the other. Other studies, however, suggest an interaction between rhythm and pitch. Deutsch's (1980) memory study found best recall of melodies in which pitch and temporal organization coincided and worst performance when they conflicted. Jones, Boltz, and Kidd (1982) also varied pitch and temporal contexts simultaneously. Their listeners were more accurate at discriminating pitch changes coinciding with important temporal events than changes coinciding with less important temporal events.

Because actual music is rich in both temporal and pitch variations, it is of considerable interest to music psychology to understand how these factors combine in perception and memory. The question is also related to a more general concern arising from the information processing tradition, namely, the capacity for encoding and processing more than one source of stimulus information simultaneously. This literature suggests two possibilities for the perception of pitch and temporal structures: perceptual independence and perceptual interaction. The former is defined formally as the case in which the effects of two stimulus components are independent random variables (Ashby & Townsend, 1986). Measures of correlation and performance parity are typically taken. If responses to two components have zero correlation across subjects and if responses to the combined components equal the sum of responses to the individual components (performance parity), then perceptual independence of the two individual components is assumed (Garner & Morton, 1969). Perceptual interaction of a unit or whole is defined operationally as the failure to attend selectively to individual parts (Pomerantz, 1981). In this case, independent processing of components forming such a unit should be difficult, whereas processing of a unit as a whole should be easy. The definition of perceptual independence encounters several problems in application, however. These include correlation of components as defined within the stimulus, lack of discrimination between perceptual and decision processes, and an interrelation between zero correlation and performance parity (Ashby & Townsend, 1986; Garner, 1981; Garner & Morton, 1969).

Music offers two main advantages for investigating the question of perceptual independence. The stimulus attributes, which in the musical case are the temporal and pitch compo-

nents, can be defined and manipulated independently. This makes possible an experimental situation in which perceptual dependence is not confounded by stimulus dependence. Thus, if an interaction is observed, it can be attributed to perceptual, rather than stimulus, dependence. Second, the question of independence arises naturally from music in which temporal and pitch structures are intentionally combined. Presumably, the way in which musical traditions have chosen to employ these structures reflects intuitions about the listener's cognitive processes. It is a generally held belief that tonally strong events tend to be placed at rhythmically stressed points in time, with the consequence that the two organizations are mutually reinforcing. If true, this suggests that listeners are limited in their capacities for processing the two organizations independently. This can be tested empirically by decoupling the two sources of information in the stimulus. Following this strategy, we present here two methods for independently varying temporal and pitch information to assess their influences on the perceptual organization of musical phrases.

In the present studies, we chose to focus on musical phrase structure as one of many forms of perceptual organization in music. This focus was motivated by a number of observations. First, a phrase is used in music-theoretic analyses to denote a unit of information or meaning (Riemann, 1896). Second, definitions of musical phrase often depend on both temporal and pitch information (Cooper & Meyer, 1960; Lerdahl & Jackendoff, 1983). Third, empirical studies provide evidence for the perceptual importance of phrase structure. Using Fodor and Bever's (1965) click-placement procedure, Gregory (1978) demonstrated that the phrase is a perceptual unit. Sloboda (1977) also found evidence for phrases when he measured the amount of musical text a reader can continue to perform after the text is removed from sight, called the eye–hand span (after Levin and Kaplan's, 1970, eye–voice span for reading text). The eye–hand span seems to stretch or shrink to phrase boundaries, demonstrating the effects of structural markers (cadences) and physical markers (spatial boundaries, such as a rest) on the readers' recall of written music. Sloboda concluded that structural markers cause the span to expand within an area of redundancy (within a phrase), whereas physical markers can cause the span to expand or contract.

Another motivation for the present study was that Lerdahl and Jackendoff's (1983) theory offers predictions for pitch and temporal structures in music. Their model is highly relevant to issues of structural independence, because it represents a listener's perception of musical structure from pitch-based and temporal-based components. The model is intended as a description of the experienced listener's musical intuitions. Explicit rules are given, some of which are based on Gestalt grouping principles. The model begins with the perceived sequence of pitches and durations, and the experienced listener applies powerful resources of expectancy to infer various kinds of hierarchical structure. The model is not completely algorithmic because it does not specify the order in which the rules composing the analysis are applied or how they are to be weighted and combined. Nevertheless, comparison of the model's predictions with the following experiments is valuable because the components' predictive power may clarify constraints on the rules

(such as order and independent/interactive application) that reflect perceived relation between pitch and temporal structures.

The theory assumes four kinds of hierarchical structure—namely, grouping structure, metrical structure, time-span reduction, and prolongational reduction. Grouping structure describes how pitch information is organized into successively larger units: from motives to phrases and to sections. Metrical structure is the interpretation of a periodic hierarchy of weak and strong beats. Time-span reduction combines grouping and metrical structures into time-spans, assigning each event to a hierarchy of structural importance. Finally, prolongational reduction assigns a hierarchy of harmonic and melodic tension and relaxation. The metrical structure and time-span reduction are of greatest relevance to the present study; the grouping structure and prolongation reduction are not relevant here because they describe structural units larger than the stimulus materials used in the present experiments.

The metrical structure and time-span reduction of Lerdahl and Jackendoff's (1983) theory will be described in more detail with reference to Figure 1, which shows the musical segment used in the present experiments. The segment consists of the first four measures of the Fugue XX in A minor by Bach from the *Well-Tempered Clavier* (Book I). This melody is a single-voice fugue subject and was chosen for its representativeness of a particular Western style of music (baroque), its complexity, and its reasonable length. A detailed analysis of this segment was supplied to us by F. Lerdahl (personal communication, May, 1985), following the model (Lerdahl & Jackendoff, 1983) discussed below.

The metrical structure assigns each beat a level of relative stress. The predicted metrical structure for the Bach fugue subject is represented by the dots under each note in Figure 1, Panel B. The numbers beneath the dots in Figure 1, Panel C, reflect the quantified metrical predictions for each musical event. The most important events in Figure 1, Panel B, are the first and third beats of each measure, implying a $\frac{2}{4}$ time signature. This metrical assignment reflects the equal probability with which Bach's fugue subjects begin on the first or third beat of a four-beat measure. (This is in contrast to the usual hierarchical assignment for four-beat measures, in which the first beat is more important than the third, and exemplifies a nonalgorithmic aspect of the model.)

The time-span reduction reflects the combined contributions of pitch and temporal hierarchies for each event relative to the whole musical segment. It represents the relative importance of each event in a nested tree, usually with binary branchings. Each branching contains a most important event called the *head*, represented with a continuing line, and a subordinate event attached to it with a noncontinuing line. The nodes (where the lines meet) indicate structural levels of relative importance, which are nested hierarchically so that the root, or top of Figure 1, is the highest level. The ordering of nodes and branches is determined by a combination of grouping and metrical rules that has both independent and interactive components (see Lerdahl & Jackendoff, 1983, chapter 6). The degree to which the structures interact differs for each musical composition. Despite this lack of confirmatory power for tests of independence, the time-span reduction does provide a quantifiable (if not completely algorithmic) distance metric for each musical
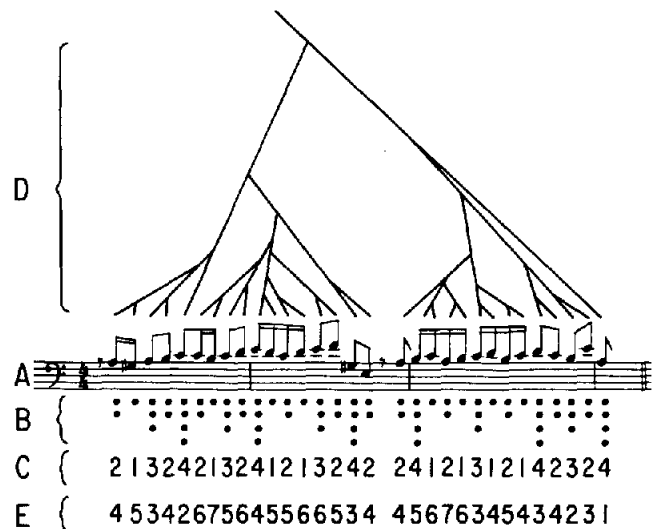

Figure 1. Metrical structure and time-span reduction tree based on Lerdahl and Jackendoff (1983) for fugue subject. Panel A: Original melody. Panel B: Dots representing levels of metrical structure for each event. Panel C: Quantified metrical structure for each event. Panel D: Time-span reduction tree. Panel E: Quantified values for time-span reduction tree. (Values indicate number of branched nodes between stem and root [hierarchical level].)

event. The time-span reduction for the Bach fugue is shown in Figure 1, Panel D, as the tree drawn above the notes. The predictions are quantified in Figure 1, Panel E, by the number of nodes branched through to connect each note with the root.

Experiment 1: Adding Patterns

In the first experiment we investigate whether or not judgments of musical phrases can be predicted from the separated pitch and temporal information. Segments were created from the opening statement of a theme from a fugue by J. S. Bach (Figure 1). Three types of segments were produced. Segments in the *pitch condition* preserved the original pitch pattern, but all tones had the same duration (equitemporal); segments in the *temporal condition* preserved the original temporal pattern, but all tones had the same pitch (equitonal); and segments in the *melody condition* retained both the original pitch and temporal patterns. Figure 2 illustrates the construction of these conditions. Trials were generated by altering the length of the excerpt so that each trial ended on a different pitch in the pitch condition, a different duration in the temporal condition, and a different combination of pitch and temporal units in the melody condition. Listeners rated how good or complete a phrase each segment made.

The psychological completeness of each unfinished segment, measured in this way, may clarify the nature of internal representations of the musical structure. If perceived goodness of a phrase is a function of pitch and temporal information, then ratings in the melody condition should be predicted accurately from a combination of the ratings from the pitch and temporal conditions. To test this, the ratings from the three conditions will be analyzed by multiple regression, predicting the melody

Figure 2. Sample trials from Experiment 1, containing complete excerpt. (Three conditions are temporal condition [Panel A], pitch condition [Panel B], and melody condition [Panel C].)

condition ratings from the other two types of ratings. The function that best predicts melodic phrase judgments from pitch and temporal factors will indicate whether or not the contributions are independent. The degree of association between pitch and temporal factors will be determined by the simple correlations between each pair of conditions. Finally, effects of phrase boundaries will be determined by examining the judgments in terms of each segment's final pitch and temporal values. Specifically, previously obtained ratings of tonal hierarchies (Krumhansl & Kessler, 1982) and Lerdahl and Jackendoff's (1983) predictions for metrical stress and time-span reduction will be compared with the experimental results. Listeners with musical backgrounds are used in order to ensure sufficient familiarity with Western music for its hierarchical organization to be revealed and for purposes of comparison with these predictive models.

Method

Subjects. Sixteen listeners from the Cornell University community were recruited. Each had a minimum of 2 years' formal training in music on a musical instrument or voice, although the average number of years' instruction was much greater (7.2 years), and the range was large (from 2.5 to 14 years). The average amount of experience (other than instruction) on an instrument or voice was 10.8 years, with a range of 3.5 to 19 years. Listeners ranged in age from 18 to 25 years (mean age = 20.1 years), and none had hearing problems or absolute pitch. Each reported being unfamiliar with the music used in the experiment and was paid $3.00 for participating.

Apparatus and stimulus materials. All stimuli were generated by a DMX-1000 digital music synthesizer (Digital Music Systems) under the control of a PDP-11/23+ computer (Digital Equipment Corporation). Signals were played over an NAD stereo 3125 amplifier through Mission Electronic loudspeakers. All tones had equal amplitudes and were played at comfortable listening levels.

Each tone contained the same harmonically complex structure: Four overtones were present above the fundamental, with successive amplitudes relative to the fundamental of 1/8, 1/4, 1/8, and 1/10. Each tone had a 20-ms rise time, with a linear decay (back to zero amplitude) over the remainder of the tone.

All stimuli were derived from the first four measures of the Fugue XX in A minor from the *Well-Tempered Clavier* (Book I), by J. S. Bach

(Figure 1, Panel A). Three conditions were created by presenting only the pitch pattern (pitch condition), only the temporal pattern (temporal condition), or both pitch and temporal patterns (melody condition) from the original fugue. In the pitch condition, all segments were equitemporal; each tone lasted 325 ms. The durations in the other two conditions were determined so that trials in all three conditions were of the same average duration. All segments in the temporal condition were equitonal (each duration was sounded on the same tone, G below middle C: 196 Hz, chosen because it is in the range of, but does not belong to, the set of pitches in the original melody), and each sixteenth note, the shortest duration, lasted 210 ms (71 beats per minute, where one quarter-note = one beat). Segments in the melody condition contained both the original temporal and pitch patterns and each sixteenth note lasted 210 ms.

Trials were generated by altering the length, which varied between units 19 (the first unit after the rest) and 32 (the last unit). Thus, each trial ended on a different pitch, or temporal event, or both, in the range of the 19th unit to the end of the chosen phrase. Because there were 14 possible ending points, 14 trials were created for each of the three conditions. Three repetitions were run, totaling 126 trials.

Procedure. Listeners were told they would hear short segments of music and were asked to rate, on a 7-point scale, how good or how complete a phrase each of the segments made (1 = a poor or incomplete phrase; 7 = a good or complete phrase). Trials consisted of the sounding of a sequence, followed by an 8-s pause, during which listeners made their ratings. Trials from the three conditions were blocked, and each block was preceded by 10 practice trials. Half of the listeners heard the melody condition first, and half heard it last. Presentation order of the temporal and pitch conditions was counterbalanced. Breaks were taken every 15 min. Listeners were run in groups of 1 to 3, and a different random sequence was created for each group. The experiment lasted 60–80 min and was run over 2 days.

Results

Mean ratings across listeners for each segment are shown in Figure 3 for melody, pitch, and temporal conditions. The horizontal axis represents the unit on which the segments ended, and the vertical axis represents the mean rating. To assess the extent of agreement among listeners, the data from each listener were correlated with those of every other listener. The mean intersubject correlation was .42 ($p < .05$), indicating significant agreement across listeners.

A 3 (condition) × 14 (ending unit) × 16 (listeners) analysis of variance was performed on the ratings. There was a significant main effect of ending unit, $F(13, 195) = 19.67$, $p < .001$. The original passage (as Bach composed it) received the highest rating in the pitch and melody conditions, and one of the highest ratings in the temporal condition. Many segments were considered to be good phrases and none of the segments received a low rating of 1. There was no significant main effect of experimental condition, but there was a Condition × Ending Unit interaction, $F(26, 390) = 7.31$, $p < .001$, indicating different effects of pitch and temporal information.

To investigate this difference, simple correlations were performed between ratings from the three conditions. The correlation between melody and temporal conditions was significant ($r = .54$, $p < .05$) as was the correlation between melody and pitch conditions ($r = .89$, $p < .05$). The correlation between pitch and temporal conditions was, however, not significant ($r = .27$, $p > .15$), indicating that strong pitch and temporal events
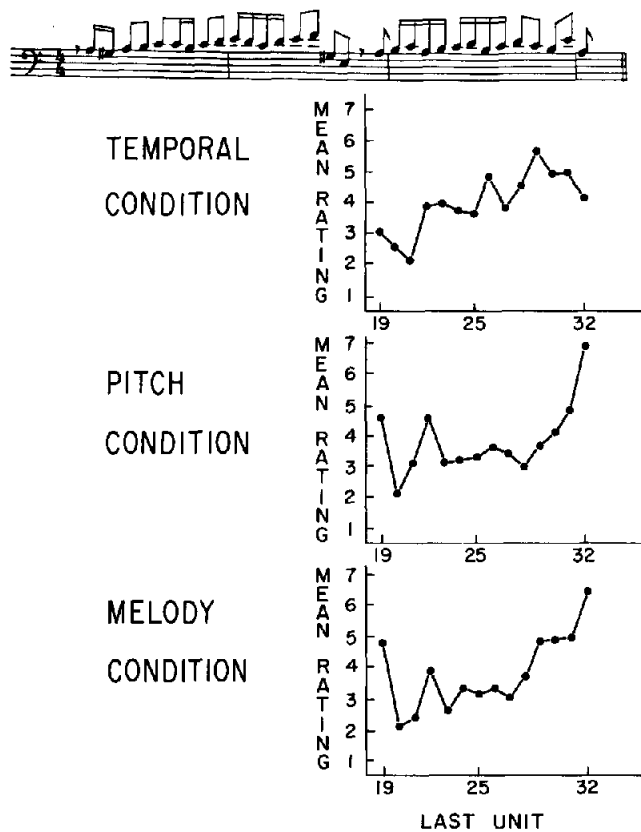
Figure 3. Mean ratings across listeners from Experiment 1 for each condition and ending unit. (Horizontal axis represents unit on which the trial ended; vertical axis represents phrase completeness ratings.)

do not coincide, but that each is associated with melody condition judgments.

A regression analysis was performed, predicting the melody condition data from the pitch and temporal condition data. A linear model provided an excellent fit to the melody condition data ($R = .94$, $p < .001$), as shown in Figure 4. Both regression coefficients were significant (standardized temporal coefficient = .32, $p < .02$; standardized pitch coefficient = .80, $p < .001$), and they did not differ significantly, $F(1, 12) = 2.05$, $p > .10$. The fit of the linear model was significantly better than the simple correlation of the melody condition with the pitch condition, $F(1, 11) = 9.1$, $p < .05$, or with the temporal condition, $F(1, 11) = 57.3$, $p < .05$. The accurate predictions of the linear regression model provide evidence for independence of pitch and temporal factors in melodic phrase judgments.

The additive combination of pitch and temporal condition judgments in predicting melody judgments was compared with several interaction models, including the multiplicative model with an added Temporal × Pitch interaction term ($M = T + P + T \times P$). A multiple regression analysis was run for this model; there was no improvement in the fit ($R = .94$, $p < .001$), and none of the regression coefficients were significant (smallest p > .20). Several other models with interaction terms were tested; none produced a better fit than the simple additive

model. This is not surprising because so little variation was unaccounted for by the additive model.

To test the additive model for individuals, regression analyses were run on each listener's data separately. The simple linear regression equation provided a good fit ($p < .05$) for 13 of the 16 listeners (the range of the multiple correlation coefficient for the 16 listeners was from .47 to .97; the median multiple correlation was .81). The simple additive model was contrasted with various interaction models. The multiplicative interaction model, along with other interaction models, did not fit the individual data better than the simple additive model.

Analyses were performed to check for effects of musical background. There were no significant correlations between amount of musical training and pitch weights ($r = .13$, $p > .60$) or temporal weights ($r = .32$, $p > .20$) in the multiple regression analyses. There was also no association of experience (performance other than formal training) with pitch weights ($r = .01$, $p > .90$) or with temporal weights ($r = .16$, $p > .50$). Nor were there any differences in the individual multiple correlations (degree of fit of the linear model) associated with musical background.

Analyses were also performed to check for effects of presentation order of the experimental conditions. There were no effects of order of pitch and temporal conditions on either pitch or temporal weights, but there was a significant effect of order of melody condition, $t(14) = 3.88$, $p < .01$, on temporal weights. Listeners who heard the melody condition first had higher standardized temporal weights (mean temporal weight = .62) in the multiple regression of melody judgments than those who heard the melody condition last (mean temporal weight = .17). There was no effect of condition order on pitch weights. The individual multiple correlation coefficients (fit of the linear model) were larger for listeners who heard the melody condition first, although they were significant for listeners in both groups. The individual correlations between pitch and temporal conditions, which were low, did not differ by melody condition order,
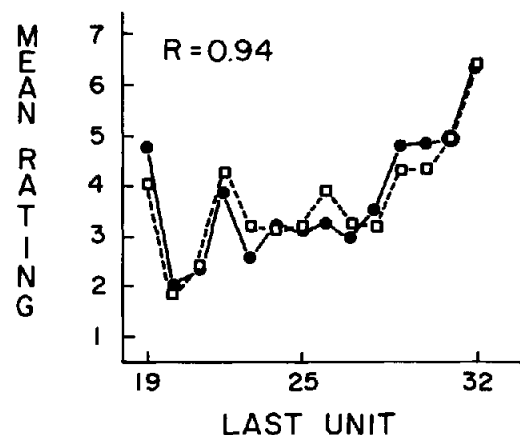

Figure 4. Fit of linear regression model to melody condition data from Experiment 1. (Horizontal axis represents unit on which trial ended; vertical axis represents mean rating of phrase completeness. Observed values from melody condition are shown in filled circles; predicted values from linear regression of temporal and pitch conditions are shown in open squares.)

$t(14) = 1.41, p > 1$. So, despite the order effect of melody condition, the simple additive model is still appropriate.

The effect of the last pitch was examined by comparing judgments from the melody and pitch conditions with ratings from previously obtained measures of tonal hierarchies. Krumhansl and Kessler's (1982) ratings of fit for pitches from the minor key were also judged on a 7-point scale, where 7 represented a very good fit of a pitch to a key-defining context. Ratings from the pitch condition in the present study correlated significantly with the key-context data $(r = .61, p < .05)$, and the correlation with melody condition ratings approached significance $(r = .47, p < .10)$.

To determine whether there were similar effects of the last beat, a 4-point scale quantifying Lerdahl and Jackendoff's (1983) metrical structure was used (as shown in Figure 1, Panel C). Segments ending on the first or ninth sixteenth-note of a measure were assigned the value 4; segments ending on sixteenth-notes 5 or 13 were assigned 3; segments ending on sixteenth-notes 3, 7, 11, or 15 were assigned 2; and all other, less important sixteenth-notes were assigned 1. These ratings did not correlate significantly with the mean ratings from the temporal condition $(r = .13, p > .60)$ or the melody condition $(r = .35, p > .20)$.

Melody condition ratings were compared with the time-span reduction, which combines both pitch and temporal information. Each event in the time-span reduction tree (see Figure 1, Panel D) is coded with a level of strength, corresponding to the number of nodes branched through on the tree. These codings, ranging from 1 (closest to root) to 7 (farthest from root), were compared with the phrase judgments. The codings were correlated with ratings from the melody condition. The correlation was significant $(r = -.68, p < .01)$, indicating that tones closest to the root in the hierarchy corresponded to the highest ratings in the melody condition. To ensure predictive power of the time-span reduction apart from the metrical structure, the melody condition ratings were correlated with the time-span reduction, after metrical structure effects were partialed out. The semi-partial correlation was significant $(r = -.58, p < .05)$, indicating the time-span reduction does contribute beyond the metrical structure.

In order to compare the independence of the sets of predictions, the quantified tonal hierarchy (Krumhansl & Kessler, 1982) was correlated with the metrical structure (Lerdahl & Jackendoff, 1983). The two sets of predictions, paired according to the last pitch and beat in each segment of the melody, did not correlate significantly $(r = -.06, p > .80)$. The metrical structure and time-span reduction (Lerdahl & Jackendoff, 1983) correlated significantly $(r = -.55, p < .05)$, although the ratings of fit for pitches (Krumhansl & Kessler, 1982) did not correlate with the time-span reduction $(r = -.02, p > .90)$. This was not surprising because the time-span reduction is based on metrical and grouping rules, which depend on a particular context, unlike the tonal hierarchy ratings, which are derived from abstract scale contexts.

Discussion

This experiment demonstrated that judgments of melodic phrases can be described in terms of phrase ratings based on pitch and temporal patterns in a Bach fugue subject. In particular, an additive model (without an interaction term) was sufficient, as shown by accurate prediction of melody condition judgments from pitch and temporal condition judgments and by favorable comparison with models containing interaction terms. This result suggests independent pitch and temporal factors in the melodic phrase judgments. These points were evidenced both in the averaged and individual data. The original segment (as Bach composed it), ending on a strong pitch and a strong temporal event, was considered the best phrase. Additionally, ratings in the pitch and temporal conditions did not correlate; nor did the metrical structure (Lerdahl & Jackendoff, 1983) and tonal hierarchy (Krumhansl & Kessler, 1982) correlate, as applied to the Bach fugue subject. These two results indicate that perceptually strong tonal events are not paired with metrically strong beats in the fugue subject.

The relative weighting of temporal information differed across listeners with the order in which the conditions were presented. Listeners who heard the melody condition first tended to show a greater correspondence between judgments in the temporal and melody conditions than those who heard the temporal or pitch condition first. Thus, there appeared to be carryover effects of the melody condition to the temporal condition but not to the pitch condition, possibly due to a stronger memory component for the pitch structure. This finding did not affect the main experimental conclusions because the correlation between pitch and temporal conditions remained low and the simple additive model fit well, regardless of the order of experimental conditions.

Finally, the influence of the last sounded event was tested for both pitch and temporal information. Predictions from the Krumhansl and Kessler (1982) data for tonal hierarchies fit the melody and pitch condition data well, corroborating the tonal hierarchy in music containing changing temporal patterns. Predictions from the metrical structure of Lerdahl and Jackendoff (1983), however, did not correlate well with ratings from the temporal condition; their metrical predictions may reflect a larger effect of pitch information than the model indicates. Another possibility is that the shortened segments did not provide enough information to establish the metrical structure of the complete excerpt. Alternatively, the influence of temporal duration patterns may be more complex than the periodic metrical structure. The time-span reduction predictions of their model (which combine pitch and temporal effects) did supply accurate predictions for melodic phrase judgments, though.

Experiment 2: Shifting Patterns

A second experiment was designed, using a different methodology, to further test the contributions of temporal and pitch factors to melodic phrase judgments. One might argue that the paradigm of the first experiment, presenting only equitonal or only equitemporal patterns, produced segments that do not resemble actual music. Two of the conditions in the first experiment contained just pitch or just temporal variations, which may have produced an artificial dissociation between the two types of information. Another possible criticism of the first experiment is that the final sounded events contributed heavily to the phrase judgments because, within condition, the beginning

Figure 5. Sample trials for Experiment 2, showing shift of one unit. (Three conditions are temporal shift condition [Panel A], pitch shift condition [Panel B], and melody shift condition [Panel C]. Arrows indicate beginning of original pattern of temporal [T1] and pitch [P1] pattern.)

of each trial was always the same and only the end of each trial differed.

To address these issues, a second experiment was run, using segments that all contained variations in both pitch and temporal information. The musical segments were based on the same excerpt in order to permit comparisons with the first experiment. Three types of segments were created, each containing both the original pitch and temporal patterns, in an altered combination. The two patterns were separated; one of the patterns underwent a phase shift and was then recombined with the unaltered framework of the other pattern. An example of the construction of these conditions for a shift of one unit to the right is shown in Figure 5. Segments in the *pitch shift condition* contained the original temporal pattern, while the pitch pattern was shifted by varying amounts (one position to the right in the example of Figure 5), wrapping the remaining pitches around to the beginning, and recombining with the temporal pattern. The analogous change was made for segments in the *temporal shift condition*, keeping the pitch pattern constant while shifting the temporal pattern. Segments in the *melody shift condition* were created by shifting the pitch and temporal patterns by the same amount and in the same direction (essentially choosing a different beginning and ending point). For every segment in the melody shift condition, there was a segment with the same (shifted) temporal pattern and original pitch pattern in the temporal shift condition, and a segment with the same (shifted) pitch pattern and original temporal pattern in the pitch shift condition. Again, listeners with musical backgrounds rated how good or complete a phrase each segment made.

As in the first experiment, a multiple regression analysis on the ratings for the three conditions will indicate whether perceived melodic phrases are a function of pitch and temporal components. If the simple additive model can predict phrase judgments from combinations of pitch and temporal information, then melody shift judgments should be fit well by a lin-

ear combination of the corresponding pitch shift and temporal shift judgments. The relative weighting of pitch and temporal information may differ from Experiment 1 because all segments in this experiment contain rich patterns of both types of information. Again, both additive and interactive models will be compared in averaged and individual data, and effects of musical background will be examined. If the correlation between the pitch shift and temporal shift conditions is low, then this provides converging evidence that temporal and pitch hierarchies do not coincide in this fugue subject. Finally, the pitch and temporal values at phrase boundaries may again explain a large amount of variance in judgments.

Method

Subjects. Sixteen listeners were recruited from the Cornell University community. They were required to have had a minimum of 2 years' formal training in Western music on an instrument or voice (average = 8.6 years, and the range was from 3 to 18 years). Listeners were between 19 and 32 years old (average age = 22 years). The average amount of experience (other than instruction) on an instrument was 7.8 years, and the range was 3 to 13 years. None had hearing problems or absolute pitch. None of the listeners had been in the first experiment, and all reported being unfamiliar with the music used in the experiment. Each either received credit toward an introductory psychology course or was paid $3.00 for participating in the experiment.

Apparatus and stimulus materials. Stimuli were generated using the same equipment as in the first experiment, except for the amplifier and loudspeakers; a Stereoscopic amplifier and KLH loudspeakers were used. Each tone contained the same harmonically complex structure as in Experiment 1.

. All stimuli were derived from the same musical passage as in the first experiment. Three conditions were created, by shifting either the pitch information (pitch shift condition), the temporal information (temporal shift condition), or both pitch and temporal information (melody shift condition). In the pitch shift condition all segments began on the first temporal unit (notated [T1]) and ended on the last temporal unit [T32] while the pitch pattern was shifted and recombined with the temporal pattern. All segments in the temporal shift condition began on the first pitch [P1] and ended on the last pitch [P32], while the temporal pattern was shifted and recombined with the pitch pattern. The melody shift condition was created by shifting both the pitch and temporal information by the same amount and in the same direction, wrapping around to the first unit, [T1, P1], after reaching the last unit, [T32, P32].

All note durations were the same as in the melody condition of the first experiment. As there were 31 events or units (excluding Unit 18, the rest, because the segments ending on the rest and before the rest are experientially the same), 31 sequences were created for each condition, giving a total of 93 trials.

Procedure. The procedure was the same as in the first experiment. Trials again consisted of the sounding of a sequence, followed by an 8-s pause, during which listeners rated how good or complete a phrase the passage made, on a scale of 1 to 7 (1 = a bad or incomplete phrase, 7 = a good or complete phrase). The experiment was run over a 2-day period. Half of the listeners performed the melody shift condition on Day 1, and the other half on Day 2. Trials from the pitch and temporal shift conditions, heard on the alternate day, were randomly intermixed.

Results

The mean rating across listeners for each segment is shown in Figure 6. The horizontal axis represents the unit (from the original segment) on which the shifted pattern ended, and the
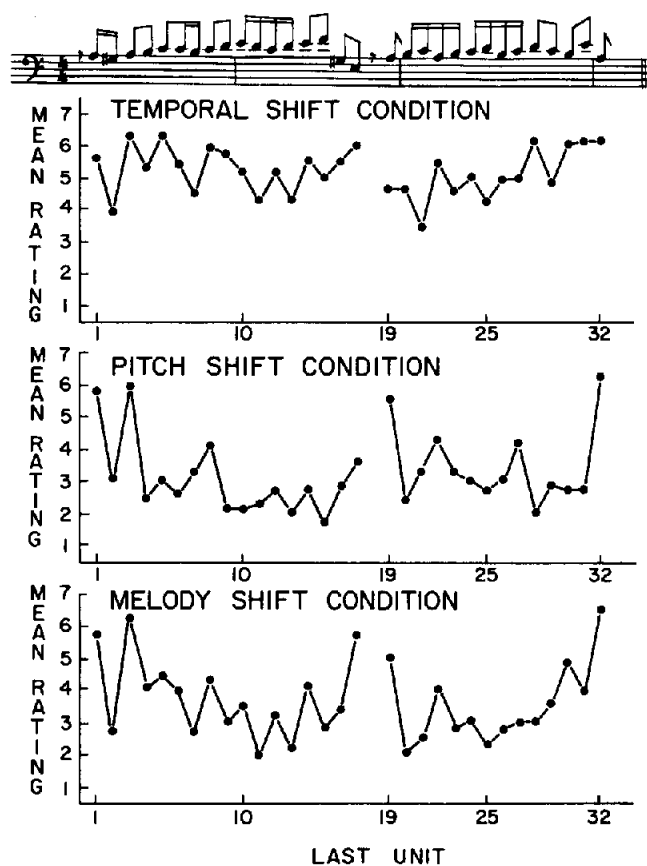
Figure 6. Mean ratings across listeners from Experiment 2 for each condition and ending unit. (Horizontal axis represents unit on which the trial ended; vertical axis represents mean phrase completeness ratings.)

vertical axis represents the mean rating. The mean intersubject correlation was .48 ($p < .05$), again indicating substantial agreement among listeners.

A 3 (condition) × 31 (ending unit) × 16 (listener) analysis of variance was performed on the phrase ratings. There was a significant main effect of ending unit, $F(30, 450) = 20.2, p < .001$, and a main effect of condition, $F(2, 30) = 40.4, p < .001$. Segments in the temporal shift condition received higher ratings on average than segments in the other two conditions. The highest mean rating in the pitch and melody shift conditions, and one of the highest ratings in the temporal condition, was again given for the original segment (as Bach composed it). There was a significant interaction of condition and ending unit, $F(60, 900) = 6.0, p < .001$, indicating different effects of temporal and pitch information.

This difference was first examined with simple correlations between ratings from the three conditions. The correlations of melody shift condition with temporal shift condition ($r = .71$, $p < .01$) and with pitch shift condition ($r = .73, p < .01$) are high. The correlation between temporal and pitch shift conditions, however, is low ($r = .24, p > .30$), indicating separate factors contributing to ratings from the melody shift condition. A regression analysis was then performed, predicting the melody shift data from the pitch shift and temporal shift data. The sim-

ple additive model provided a good fit, as shown in Figure 7 ($R = .91, p < .001$). Both regression coefficients were significant (temporal standardized coefficient = .56, $p < .001$; pitch standardized coefficient = .59, $p < .001$), and they did not differ significantly, $F(1, 29) = 2.27, p > .13$. The fit of the simple additive model was better than either of the simple correlations of melody shift data with pitch shift, $F(1, 28) = 50.8, p < .05$, or with temporal shift data, $F(1, 28) = 56.1, p < .05$. Again, the adequacy of the simple additive model indicates independent pitch and temporal factors in melodic phrase judgments.

The simple additive model was again contrasted with several models assuming interaction between pitch and temporal factors. The simplest multiplicative interaction model ($M = T + P + T \times P$) did not fit better ($R = .91, p < .001$) than the simple additive model, and none of the regression coefficients were significant. None of the other complex models with interaction terms produced a better fit than the simple additive model.

To examine the fit of the additive model for individuals, regression analyses were run on each listener's data. The simple additive model provided a significant fit ($p < .05$) for 12 of the 16 listeners; the multiple correlation coefficients for the 16 listeners ranged from .15 to .82, and the median value was $R = .53$. For each listener, this model was contrasted with models containing an interaction term. None of the interaction models provided a better fit than did the simple additive model.

Effects of musical background and order of experimental conditions were examined. There were no significant correlations between the individual temporal or pitch correlation weights with musical background, and no differences due to the order of experimental conditions. Nor were there any differences in the multiple correlation coefficients corresponding to musical background or order of experimental conditions.

To examine effects of the final pitch and temporal events on judgments, analyses were again performed comparing the last unit in each segment with previously obtained results. Krumhansl and Kessler's (1982) tonal hierarchy for minor keys was correlated with the phrase judgments in the pitch shift condition on the basis of the last pitch of each segment. Their data correlated significantly with the pitch shift condition ($r = .72$, $p < .001$), and the melody shift condition ($r = .66, p < .001$), again indicating that the tonal hierarchy operates in music containing complex temporal patterns.

The effects of the last temporal unit were tested using the same four-level metrical hierarchy as before from Lerdahl and Jackendoff (1983) (see Figure 1, Panel C). This scale correlated significantly with both the temporal shift condition ($r = .66$, $p < .001$) and the melody shift condition ($r = .39, p < .05$), a result that differs from the first experiment.

The time-span reduction tree of Figure 1, Panel C, from the model of Lerdahl and Jackendoff (1983) was compared with the melody shift ratings. The quantified predictions, ranging from 1 (closest to root) to 7 (farthest from root) correlated significantly with the melody shift ratings ($r = -.52, p < .01$), indicating that important tones in the hierarchy corresponded with high ratings in the melody shift condition. Again, the semi-partial correlation between melody shift ratings and time-span reduction was calculated after metrical structure effects were removed. The correlation was significant ($r = -.36, p < .05$), indi-
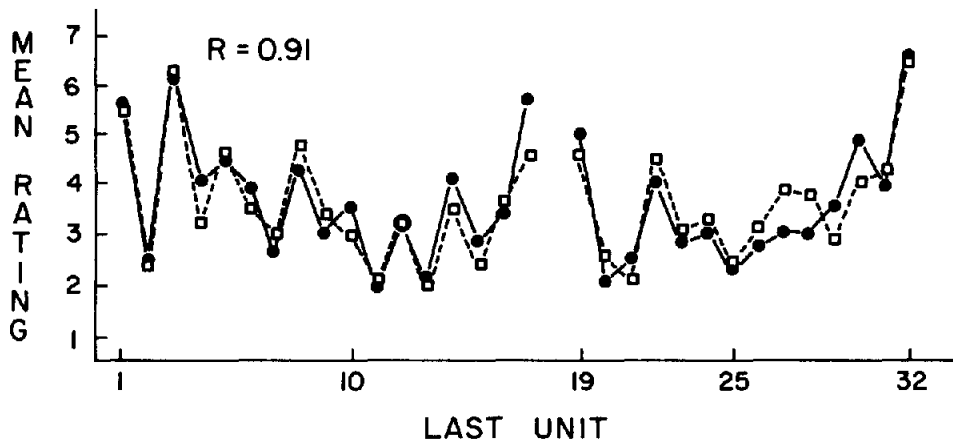
Figure 7. Fit of linear regression model to melody shift condition data from Experiment 2. (Horizontal axis represents unit on which the shifted pattern ended; vertical axis represents mean phrase completeness ratings. Observed values from melody shift condition are shown in filled circles; predicted values from linear regression of temporal and pitch shift conditions are shown in open squares.)

cating that the time-span reduction has predictive power apart from the metrical structure.

The three sets of predictions were again compared. Krumhansl and Kessler's (1982) ratings of fit for pitches did not correlate significantly with the metrical structure predictions of Lerdahl and Jackendoff (1983) for the original melody ($r = .16, p = .40$). The metrical structure and time-span reductions correlated significantly ($r = -.59, p < .01$), whereas the ratings of fit for pitches did not correlate with the time-span reduction ($r = -.19, p > .20$). Again, this reflects the formation of the time-span reduction on metrical and grouping structures largely dependent on a particular context.

Comparisons were made between the first and second experiments. A correlation was performed on the average ratings for each segment ending on the same unit within stimulus condition. The correlation of the melody condition from Experiment 1 with the melody shift condition from Experiment 2, for segments ending on units 19 to 32, was highly significant ($r = .92, p < .001$). This was expected to be high, because the only difference between these stimuli was the amount of preceding context. The correlation of the temporal condition from Experiment 1 with the temporal shift condition of Experiment 2 ($r = .61, p < .02$) and the correlation of the pitch and the pitch shift conditions ($r = .77, p < .01$) were both significant, indicating significant agreement between the two experiments.

Discussion

This experiment verified with a different methodology the previous finding that phrase judgments based on pitch information and temporal information are sufficient to predict judgments of melodic phrases in a Bach fugue subject. As in the first experiment, the simple additive model, predicting melody shift ratings from temporal shift and pitch shift ratings, was sufficient; models with interaction terms provided no better fits. This result demonstrates independent contributions of pitch and temporal factors to melodic phrase judgments in the ex-

cerpt. These points were confirmed both in the averaged and the individual data. Again, Bach's original composition received the highest rating. There was significant agreement between corresponding conditions in the two experiments, suggesting that the absence in Experiment 1 or presence in Experiment 2 of one original pattern of information did not significantly alter judgments made on the other pattern.

The relative weighting of pitch and temporal factors did not differ across listeners as a function of the order in which they received the experimental conditions. This finding differed from the first experiment, possibly because of the more complex information in the stimulus materials (both temporal and pitch variations were present in every trial), creating richer, more musical materials in each condition. Alternatively, the length of the average segment increased in the second experiment, possibly providing more information for phrase judgments.

Again, the temporal and pitch conditions did not correlate, indicating that the two perceptual hierarchies of rhythm and tonality do not coincide in the fugue subject. The phrase boundaries explained a significant amount, but not all, of the variance in judgments. The ratings of tonal hierarchies from Krumhansl and Kessler (1982) correlated well with the experimental results, proving robust across music with complex temporal patterns. The metrical structure from Lerdahl and Jackendoff (1983) was also supported by the experimental results. This corroboration, contrasting with the first experiment, may be due to the longer average trial length (segments in the first experiment had an average length of 25 units, whereas every segment in the second experiment contained 32 units), allowing the metrical structure to be more firmly established in this experiment.

The time-span predictions, although intended only for the original combination of pitch and temporal patterns and not for the shifted segments, also corresponded to shifts in starting and ending points in the melody shift condition. In music with larger pitch shifts or less regular temporal patterns than the fugue segment used here, shifted patterns might produce

.greater incoherence, and hence a time-span analysis of shifted patterns might be inappropriate. Alternatively, the time-span reduction may fit shifted segments well due to a precedence of relations between adjacent events (groups formed early) over nonadjacent events (groups formed later), supporting a bottom-up processing model. The rules that form the time-span analysis reflect both types of relations and the order in which the rules operate have not yet been determined. However, the present results seem to support an additive rather than interactive combination of rules in formation of time-span analysis.

General Discussion

The present findings indicate that melodic phrase judgments in a Bach fugue subject are based on two independent types of structure: temporal and pitch information. This conclusion was supported by experiments using two different paradigms, each of which preserved one pattern of information while altering the other. The effects of pitch and temporal events were additively combined with fairly equal weights to predict phrase judgments in this baroque melody. There was very little association between the two types of structures comprising melodic phrases, and models assuming interaction between them provided no better a fit than did the additive model. There did not appear to be any emergent perceptual qualities of musical phrases containing both pitch and temporal patterns that were not found in the ratings of the separately presented components. Furthermore, there were no differences in the components' weights associated with musical training or experience, indicating a large range of listeners for whom temporal and pitch hierarchies operate independently in phrase determination.

Temporal and pitch information may engage independent mental processes due to internal constraints on the way we produce and perceive information across temporal frames. There are variations across performances in production of timed intervals (from sources such as neural limits on the precision of motor programming). If there were an interaction between temporal and pitch factors, every performance of the same piece of music might provide a very different perception. Recognition of familiar melodies is quite robust, however, when certain changes are made in one component. For example, a change in tempo (speeding up and slowing down the temporal rate) or a transposition (change in absolute pitch, retaining relative pitch differences) does not change a tune's identity.

The present results satisfy information processing definitions of perceptual independence in a domain in which the components' combinations occur naturally. Perceptual independence is supported by the independent manipulation of temporal and pitch components within the stimuli, combined with low correlations between ratings of temporal and pitch components, low correlations between the metrical and tonal predictions, and the observed performance parity of temporal and pitch components in predicting melodic phrases.

The information at phrase boundaries explains a large percentage (although not all) of the variance in judgments; the proportion of variance not accounted for by the final items is presumably a function of the whole segment. The hierarchical ordering of pitch importance obtained by Krumhansl and Kessler

(1982) correlated highly with the ratings as a function of the last sounded pitch. This perceived structure, previously measured with abstract key-defining contexts, also operates in the more complex musical conditions of the fugue segments. The phrase ratings also agreed with the metrical and time-span predictions of Lerdahl and Jackendoff (1983). However, the present results suggest two constraints on these theoretical structures: The metrical structure may be instantiated only in the presence of changing pitch patterns, and rules comprising the time-span reduction may be additive, rather than interactive, for melodies of the baroque style.

The generality of independent temporal and pitch factors should be investigated with different musical excerpts within and across styles. The independence of pitch and temporal factors in musical phrase judgments may be particularly suited to the music of Bach and other contrapuntal composers, in which identical or highly similar temporal patterns are often offset in pitch. Temporal and pitch structures may coincide to a greater degree in musical styles with less regular patterns of organization, such as Gregorian chant or free-form jazz, where meters are often aperiodic and representational hierarchies must change or remain fluid over relatively long periods of time. The relation between temporal and pitch structures may also change for music that places a greater emphasis on vertical (harmonic) organization, and explorations are currently under way to examine this possibility.

References

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93,* 154–175.

Benjamin, W. E. (1984). A theory of musical meter. *Music Perception, 1,* 355–413.

Bharucha, J. J. (1984). Anchoring effects in music: The resolution of dissonance. *Cognitive Psychology, 16,* 485–518.

Cooper, G. W., & Meyer, L. B. (1960). *The rhythmic structure of music.* Chicago: University of Chicago Press.

Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception & Psychophysics, 28,* 381–389.

Deutsch, D., & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. *Psychological Review, 88,* 503–522.

Essens, P. J., & Povel, D-J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception & Psychophysics, 37,* 1–7.

Fodor, J. A., & Bever, T. G. (1965). The psychological reality of linguistic segments. *Journal of Verbal Learning & Verbal Behavior, 4,* 414–420.

Garner, W. R. (1981). The analysis of unanalyzed perceptions. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 119–139). Hillsdale, NJ: Erlbaum.

Garner, W. R., & Morton, J. (1969). Perceptual independence: Definitions, models, and experimental paradigms. *Psychological Bulletin, 72,* 233–259.

Gregory, A. H. (1978). Perception of clicks in music. *Perception & Psychophysics, 24,* 171–174.

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review, 83,* 323–355.

Jones, M. R., Boltz, M., & Kidd, G. (1982). Controlled attending as a function of melodic and temporal context. *Perception & Psychophysics, 32,* 211–218.

Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology, 11,* 346–374.

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

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

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

Levin, H., & Kaplan, E. A. (1970). Grammatical structure and reading. In H. Levin & J. P. Williams (Eds.), *Basic studies on reading* (pp. 119–133). New York: Basic Books.

Longuet-Higgins, H. C., & Lee, C. S. (1982). The perception of musical rhythms. *Perception, 11,* 115–128.

Martin, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review, 79,* 487–509.

Monahan, C. B., & Carterette, E. C. (1985). Pitch and duration as determinants of musical space. *Music Perception, 3,* 1–32.

Pomerantz, J. R. (1981). Perceptual organization in information processing. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 141–180). Hillsdale, NJ: Erlbaum.

Povel, D-J. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 3–18.

Riemann, H. (1896). *Dictionary of music* (J. S. Shedlock, Trans.). London: Augener & Co.

Shepard, R. N. (1982). Geometric approximations to the structure of musical pitch. *Psychological Review, 89,* 305–333.

Simon, H. A., & Sumner, R. K. (1968). Pattern in music. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 219–250). New York: Wiley.

Sloboda, J. (1977). Phrase units as determinants of visual processing in music reading. *British Journal of Psychology, 68,* 117–124.

Tenney, J., & Polansky, L. (1980). Temporal Gestalt perception in music. *Journal of Music Theory, 24,* 205–241.

Psychological Documents to Resume Operation

On June 16, 1986, the on-demand publication system *Psychological Documents,* published by the American Psychological Association from 1971 through 1985, was sold to Select Press. Select Press will begin publishing new volumes this year and, as of June 16, 1986, began fulfilling orders for documents accepted into the system while it was published by APA.

Peer-reviewed documents were published by APA under the experimental system (formerly Journal Supplement Abstract Service) for 15 years. A catalog containing synopses of each document accepted into the system was published on a subscription basis. Those wishing to have a copy of the full-text of a document could order a copy in either microfiche or paper.

During periodic evaluation of the service, however, APA found that as a result of low volume, the difficulties of providing service within existing systems, the expenses related to fulfilling orders, and the cost of maintaining an editorial office, it was extremely difficult for APA to maintain service that was both timely and economical. After an extensive review of the history of the system and intensive evaluation of the expenses related to it, the APA Council of Representatives voted in 1985 to discontinue publication of *Psychological Documents* with publication of the December 1985 catalog. APA was to continue to fulfill orders for individual copies of documents until December 1986, assuming that no alternative publisher could be found. Possible alternative publishers included APA divisions, individuals, and commercial publishers. In mid-1985, Select Press approached APA and negotiations were begun.

Select Press will continue to operate the system as a peer-reviewed "journal" or document service. It will continue to feature specialized documents suitable for individual circulation such as technical reports, annotated and technical bibliographies, original data sets, test instruments, test manuals, and papers that would ordinarily be too long to be considered for regular journals. Select Press expects to expand the system to cover a broader range of documents including interdisciplinary content and possibly brief, early announcements of new findings. Select Press also publishes the interdisciplinary *Journal of Social Behavior and Personality.* Further information about *Psychological Documents* may be obtained from Select Press at P.O. Box 9838, San Rafael, California 94912.