Pitch and temporal contributions to musical phrase perception: Effects of harmony, performance timing, and familiarity

CAROLINE PALMER and CAROL L. KRUMHANSL
Cornell University, Ithaca, New York

Four experiments assessed pitch and temporal contributions to phrase judgments made on excerpts from classical music. In Experiment 1, pitch-condition trials retained the original pitch pattern but were equitemporal, temporal-condition trials retained the original temporal pattern but were equitonal, and combined-condition trials contained both patterns. In Experiment 2, one pattern was shifted in phase and recombined with the other pattern to create the pitch and temporal conditions; in the combined condition, both patterns were shifted together. The stimuli in Experiments 3 and 4 used durations from a recorded performance. In all experiments, a linear combination of the pitch- and temporal-condition ratings accurately predicted the combined-condition ratings. Familiarity with the music resulted in a higher correlation between pitch- and temporal-condition ratings, but did not alter the additive relationship; performance timing had little effect.

Listening to music involves complex auditory processing of relationships among pitches, durations, timbres, and intensities. The sound events are heard as organized into well-formed, coherent musical phrases. The experiments reported here focus on how two components of musical structure, pitch and temporal, contribute to perceived musical phrases. These components can be defined independently of one another in the music; the temporal component can be described without reference to pitch and vice versa.

This raises the question of how these components are perceived in combination. The theoretical and experimental literatures suggest at least two possible relationships; the first is a perceptual interaction. Jones (1976) describes a model in which the perceived temporal structure affects perceived pitch structure by guiding attention to the pitch events coinciding with stressed temporal events. Jones, Boltz, and Kidd (1982) found some support for this description; listeners were more accurate at discriminating pitch changes that coincided with important temporal events than changes that coincided with less important temporal events. Additional evidence for the influence of temporal structure on perceived pitch structure was found by Deutsch (1980). Memory for pitch sequences was dependent on the perceived temporal frame, such that temporal structures that coincided with pitch structures enhanced recall, whereas temporal structures that conflicted with pitch structures worsened performance.

Another possible relationship is that temporal and pitch factors are perceived independently. Monahan and Carterette (1985) found experimental support for an independent relationship in similarity ratings of melodies. In general, a primacy of rhythmic over pitch variables was found for melodies whose pitch and rhythmic elements were combined orthogonally. However, across listeners, the weights given to rhythmic and pitch dimensions were negatively correlated, suggesting that listeners perceived the dimensions separately and attended to one stimulus dimension at the expense of the other. Gabriellson's (1973) similarity studies of melodies varying in rhythm also support an independent relationship of rhythmic and pitch variables. The similarity ratings revealed the same perceived rhythmic dimensions for sequences with or without pitch variations.

In a previous study (Palmer & Krumhansl, 1987), the pitch and temporal components of a fugue subject were found to be independent and to have additive influences on judgments of musical phrases. Melodic excerpts from a baroque fugue were judged on how good a phrase they made while either temporal or pitch information was altered. These earlier findings may have resulted from the characteristics of the particular musical excerpt under study. The fugue, which employs the independent motion of individual voices, may have encouraged perceptual separation of pitch from temporal patterns. A fugue subject can be described as a temporal and melodic contour pattern that is repeated in other voices with different pitches, a form of horizontal organization.

Pitch and temporal components may not have independent influences in musical styles with primarily vertical (harmonic) organization. Music with harmonic organization, characterized by chordal progressions and thus more complex tonal development than found in single-voiced

This research was supported by a predoctoral fellowship from the National Science Foundation to the first author and by Grant MH 39079 from the National Institute of Mental Health to the second author. We are grateful to Fred Lerdahl, who generously supplied an analysis of the Mozart sonata theme. Correspondence concerning this article may be addressed to either author at the Department of Psychology, Uris Hall, Cornell University, Ithaca, NY 14853.

melodies, provides a more rigorous test of independence between pitch and temporal components. It may be, for instance, that chordal progressions impose restrictions on rhythmic progressions such that a listener's organized percept is based on an interaction of tonal and rhythmic components. Thus, the resulting percept for harmonic music may not be accounted for solely in terms of the separate components or their additive combination.

The first objective of this study was to investigate the possible influences of harmonic organization on the perception of musical phrases, using a classical harmonic excerpt. We used two methods introduced previously (Palmer & Krumhansl, 1987) for independently varying temporal and pitch information to assess their influences on musical phrase judgments, with the opening theme from the A Major piano sonata by Mozart, K. 331 (shown in Figure 1, Panel A).

A second objective was to examine the effect of listeners' familiarity with the music on the contributions of pitch and temporal components to phrase judgments. It may be that listeners familiar with a particular musical piece experience coherent, fused pitch and temporal struc-

tures, whereas the perception of listeners unfamiliar with a piece may be more dependent on grouping mechanisms that tend to operate on the separate components. To address this issue, both listeners familiar and listeners unfamiliar with the harmonic excerpt were included in the experiments reported below.

The third objective was to investigate the effect of performance timing on musical phrase judgments. In musical performance, the durations of the tones often deviate systematically from the notated durations. For example, Bengtsson and Gabrielsson (1983) reported systematic lengthening and shortening of adjacent notes, serving to alter the duration ratios between the events. Clarke (1985) assessed the effect of metrical context on piano performance deviations and found that notes in strong metrical positions tended to be lengthened and notes in weak metrical positions tended to be shortened. Many reports (Bengtsson & Gabrielsson, 1983; Shaffer, Clarke, & Todd, 1985; Todd, 1985) have demonstrated that, in performance, durations of notes completing musical phrases are lengthened relative to durations within the phrase. These deviations from mechanical regularity serve to com-

Figure 1. Metrical structure and time-span reduction tree based on Lerdahl and Jackendoff (1983) for sonata theme. (Panel A) Original theme. (Panel B) Dots representing levels of metrical structure. (Panel C) Quantified numerical assignments for metrical structure. (Panel D) Time-span reduction tree. (Panel E) Quantified values for time-span reduction tree. Values indicate number of nodes above and including first branching node, between stem and root (hierarchical level).

municate aspects of musical structure, including meter and phrasing. It may be that performance deviations influence perceived phrase structure, and thus introducing performance timing in the present study may result in a greater correlation between temporal and pitch components or an interaction in predicting the combination. To test these hypotheses, we used the durations recorded from a live performance of the excerpt in the last two experiments.

The experiments in the present study were based on two paradigms reported earlier (Palmer & Krumhansl, 1987). In both paradigms, musical segments are rated on how good or complete a phrase they made. The first paradigm addresses independence and additivity issues by presenting pitch, temporal, or the combined pitch and temporal patterns from the excerpt. If perceived goodness of a phrase is a function of pitch and temporal information, then ratings in the combined 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 were analyzed by multiple regression. The function that best predicted combined phrase judgments from pitch and temporal factors would indicate whether or not the contributions were additive. The degree of independence or association between pitch and temporal factors would be determined by the simple correlations between the two conditions.

The second paradigm addresses the same questions by separating the pitch and temporal patterns, shifting one pattern, and then recombining it with the other. In a third condition, the two patterns are shifted together. Again, a regression analysis would indicate whether each of the individual shifted patterns combined additively to predict phrase ratings for the two patterns shifted together. Effects of familiarity were examined in the correlation between temporal and pitch ratings and their contributions to the combined information. Effects of performance timing were studied in the last two experiments, in which the two paradigms were repeated with stimulus durations taken from performance measures.

Previous descriptions of tonal and rhythmic hierarchies (Krumhansl & Kessler, 1982; Lerdahl & Jackendoff, 1983) will be compared with the phrase ratings in terms of each segment's final pitch and temporal values. Krumhansl and Kessler's (1982) ratings of fit for pitches to key-

defining contexts (such as scales and chord cadences) were judged on a 7-point scale, where 7 represented a very good fit. If these previously obtained tonal hierarchies for individual pitches were to predict the phrase ratings, this would suggest that the tonal hierarchies were a component of perceived pitch organization in harmonic music.

Lerdahl and Jackendoff's (1983) theory is appropriate for comparison with the present study because it describes a listener's perception of rhythmic and harmonic, as well as melodic, organization. Their model describes four structures: grouping structure, metrical structure, time-span reduction, and prolongational reduction. Predictions from the metrical structure and time-span reduction are of greatest relevance to the present study. Metrical structure is the interpretation of a periodic hierarchy of weak and strong beats, indicated by the number of dots in Figure 1, Panel B. Time-span reduction combines metrical structures with grouping structures (such as motives, phrases, and sections) into time spans, assigning each event to a hierarchy of structural importance, represented by the tree in Figure 1, Panel D.

EXPERIMENT 1: ADDING PATTERNS

In the first experiment, three types of segments were created. Segments in the *pitch condition* preserved the original pitch pattern, but all tones had the same duration; segments in the *temporal condition* preserved the original temporal pattern, but all tones had the same pitch; and segments in the *combined condition* retained both the original pitch and the original temporal patterns. Figure 2 illustrates the construction of the pitch and temporal conditions; the combined condition is shown in Figure 1, Panel A. Trials were generated by altering the length of the excerpt so that each trial ended at a different point within the excerpt's second phrase (measures 5-8). Two groups of listeners, musicians who were either familiar or unfamiliar with the excerpt, rated how good or complete a phrase each segment made.

Method

Subjects. Sixteen adult listeners from the Cornell University community were recruited. The average amount of formal training on

Figure 2. Sample trials from Experiment 1, containing complete excerpt. Conditions shown are: (Panel A) pitch condition and (Panel B) temporal condition.

a musical instrument or voice was 7.9 years; the range was from 2.5 to 13 years. None of the listeners had hearing problems, and each was paid \$4/h for participating.

Apparatus and Stimulus materials. All stimuli were generated by a DMX-1000 signal processing computer (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: two overtones were present above the fundamental, with successive amplitudes relative to the fundamental of $\frac{1}{4}$ and $\frac{1}{100}$. Each tone had a 10-msec rise time, followed by a linear decay over the duration of the tone, to a level proportional to the total duration. The final amplitude for a duration of 1,300 msec (the longest duration in the excerpt) was 45% of peak amplitude. This waveform mimicked a 33% pulse wave with a high-level (80%) low-pass filter (the waveform on the equipment used to collect the performance measures that determined timing of stimuli in Experiments 3 and 4). Each tone had the same peak amplitude.

All stimuli were derived from the first eight measures of the sonata theme, shown in Figure 1. In the pitch condition, all tones were equitemporal with durations of 545 msec, determined so that trials in all three conditions were of the same average duration. In the temporal condition, all segments were equitonal; the pitch G above middle C (394 Hz) was chosen because it is in the range of, but does not belong to, the set of pitches in the original theme. Each sixteenth note, the shortest duration, lasted 217 msec (46 beats per minute, where one beat = three eighth notes). Each note in the original score was represented in the temporal condition by a digitally produced G with the same duration notated in the score, resulting in multiple Gs per notated chord; this method generated the same overall amplitude for trials in all three conditions. In the combined condition, segments contained both the original pitch and the original temporal patterns and each sixteenth note lasted 217 msec.

Each trial ended on a different event, anywhere from the first unit in the fifth measure to the last unit in the eighth measure. As there were 17 possible ending points, 17 trials were created for each of the three conditions. Three repetitions of each trial type were run, giving a total of 153 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-sec 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 combined 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 listeners were asked at the end of the experiment if they were familiar with the excerpt. The experiment lasted 70-90 min and was run over 2 days.

Results

Results for the two groups of listeners will be described separately in order to examine effects of familiarity on phrase ratings. Because there were no effects in any of the experiments of listeners' musical training or order of experimental conditions, responses were combined on these factors.

Unfamiliar listeners. To assess the extent of agreement among listeners, the data from each listener were correlated with the data of every other listener. The mean intersubject correlation for unfamiliar listeners was .57

Table 1
Simple Correlations in Experiment 1
as a Function of Listener Group

	Conditions			Predictions		
	Pitch	Temporal	Combined	Tonal	Metrical	Time-span
Unfamiliar Listeners						
Pitch	.44	.87†	.67†			
Temporal		.75†			.75†	
Combined			.63†	.48†		-.63†
Familiar Listeners						
Pitch	.78†	.93†	.61†			
Temporal		.90†			.64†	
Combined			.66†	.46*		-.65†

Note—All $df = 15$. * $p < .10$. † $p < .05$.

($p < .05$), indicating significant agreement within the group. Simple correlations between ratings from the three conditions (Table 1) were computed. The correlation between the pitch and combined conditions was significant, as was the correlation between the temporal and combined conditions. The correlation between pitch and temporal conditions was, however, not significant, indicating that strong pitch and temporal events do not coincide, but that each is associated with judgments in the combined condition.

A regression analysis was performed that predicted the combined-condition data from the pitch- and temporal-condition data. A linear model provided an excellent fit to the combined-condition data ($R = .96$, $p < .001$). Both regression coefficients were significant (standardized pitch coefficient = .67, $p < .001$; standardized temporal coefficient = .45, $p < .001$), and the weights did not differ significantly [$F(1,15) = 0.79$, $p > .30$]. The accurate predictions of the linear regression model, both for the grouped data and for individual subjects, provide evidence for additivity of pitch and temporal factors in phrase judgments of the excerpt containing both pitch and temporal variations. Several other models with interaction terms were compared with the additive model; none produced a better fit than the simple additive model.

The effect of previously described tonal hierarchies was examined by comparing judgments from the combined and pitch conditions with Krumhansl and Kessler's (1982) tonal hierarchy ratings for pitches from the major key. The ratings for the tones comprising the melody (highest voice) correlated significantly with both the pitch and combined conditions, as shown in Table 1.

To determine whether there were effects of the metrical predictions, a 5-point scale quantifying Lerdahl and Jackendoff's (1983) metrical structure was constructed (as shown in Figure 1, Panel C). Segments ending on the first sixteenth note of every other measure (beginning with the first measure) were assigned the value 5; segments ending on the first sixteenth note of the remaining measures were assigned 4; segments ending on the seventh sixteenth note were assigned 3; segments ending on sixteenth notes 3, 5, 9, or 11 were assigned 2; and all other, less important sixteenth notes were assigned 1. As shown

in Table 1, these ratings for the last beat in each segment correlated significantly with both the temporal-condition and combined-condition ratings.

Combined-condition ratings were compared with predictions from the time-span reduction. Each event in the time-span reduction tree was coded with a level of strength corresponding to the number of nodes passed through on the tree, above and including the first branching node (the values are shown in Figure 1, Panel E). These numbers, ranging from 1 (closest to root) to 7 (farthest from root), correlated significantly with the combined-condition ratings, as shown in Table 1. This indicates that tones closest to the root in the hierarchy corresponded to the highest ratings in the combined condition. To ensure predictive power of the time-span reduction apart from the metrical structure, the combined-condition ratings were correlated with the time-span reduction after metrical structure effects were partialled out. The semipartial correlation approached significance ($r = -.38$, $p = .13$), indicating that the time-span reduction does contribute beyond the metrical structure.

To evaluate 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 predictors, paired according to the last pitch (in the highest voice) and beat in each segment, did not correlate significantly ($r = .37$, $p > .10$). The time-span reduction correlated significantly with the metrical structure ($r = -.60$, $p < .05$) and with the tonal hierarchy ratings ($r = -.50$, $p < .05$). This was not surprising, because the time-span reduction incorporates both metrical and tonal information.

Familiar listeners. The mean intersubject correlation for familiar listeners was .66 ($p < .05$), indicating agreement within the group. Simple correlations were again performed between ratings from the three conditions (Table 1). The correlation between the pitch and combined conditions was significant, as was the correlation between temporal and combined conditions. However, the correlation between pitch and temporal conditions was also significant, indicating that pitch and temporal ratings were associated, and not independent, for familiar listeners.

A linear regression model, predicting the combined condition data from the pitch and temporal condition data, provided an excellent fit ($R = .97$, $p < .001$). Both regression coefficients were significant (standardized pitch coefficient = .59, $p < .001$; standardized temporal coefficient = .44, $p < .01$) and the weights did not differ significantly [$F(1,15) = 0.02$, $p > .80$]. The additive combination of pitch and temporal condition judgments was compared with several interaction models, both on a group and an individual basis; none produced a better fit than the simple additive model.

As also shown in Table 1, Krumhansl and Kessler's (1982) tonal profiles correlated significantly with both the pitch-condition ratings and the combined-condition ratings. The 5-point scale quantifying Lerdahl and Jackendoff's (1983) metrical structure correlated significantly

with the temporal-condition ratings, and the correlation with the combined condition approached significance. Finally, the time-span reduction predictions correlated significantly with the combined-condition ratings, indicating that tones closest to the root in the hierarchy corresponded to the highest ratings in the combined condition. The semipartial correlation of combined-condition ratings with the time-span reduction, after metrical structure effects were partialled out, approached significance ($r = -.42$, $p = .09$).

Familiarity with the musical excerpt differentially affected the simple correlation between the pitch and temporal conditions, such that familiar listeners had higher correlations than did unfamiliar listeners. The mean correlation between the pitch and temporal conditions for familiar listeners (.61) was significantly higher than that for unfamiliar listeners (.28) [$t(14) = 3.05$, $p < .01$].

Discussion

This experiment extended the results of the earlier study (Palmer & Krumhansl, 1987), in showing that phrase judgments for a sonata theme with harmonic organization can be described in terms of ratings based on the separate pitch and temporal patterns. In particular, an additive model (without an interaction term) was sufficient for both familiar and unfamiliar listeners, as shown by the accurate prediction of combined-condition judgments from pitch- and temporal-condition judgments and by the favorable comparison with models containing interaction terms. These points were evidenced in both the averaged and individual data.

As previously found, ratings in the pitch and temporal conditions did not correlate with each other for listeners unfamiliar with the excerpt. This result indicates that perceptually strong tonal events are not paired with strong metrical beats for these listeners. In contrast, the association between pitch and temporal ratings was found to be strong for listeners familiar with the excerpt. Reports of imagery for the original excerpt during the temporal and pitch conditions suggest that imagery for the missing component was responsible for the high correlation. That is, in the temporal condition, listeners may have imagined the pitches from the original music, and in the pitch condition, they may have imagined the original rhythm. Previous findings (White, 1960) have demonstrated that familiar tunes could be identified from equitonal or equitemporal patterns, supporting the idea that imagery in the temporal condition may be imposing pitch structure, and vice versa. In a paradigm that is likely to reduce this kind of imagery, then, pitch and temporal ratings should show greater independence for all listeners. The second experiment tested this prediction.

Finally, the influence of previously described tonal and rhythmic hierarchies on phrase judgments was demonstrated for all listeners. Predictions from the tonal hierarchy for the melody line (Krumhansl & Kessler, 1982) fit the combined- and pitch-condition data well, indicating that the tonal hierarchy is operative in music contain-

ing multiple voices. Predictions from the metrical and time-span structures of Lerdahl and Jackendoff (1983) fit the temporal and combined conditions, respectively. These results suggest that the prior descriptions of tonal and rhythmic hierarchies contribute to phrase judgments of harmonic as well as melodic (Palmer & Krumhansl, 1987) musical excerpts.

EXPERIMENT 2: SHIFTING PATTERNS

The second experiment was conducted in part to address the possible role of imagery in the first experiment and, more generally, to examine the effects of tonal and rhythmic structures, using a methodology that separates these two components in a different way. The impoverished segments of the first experiment (those containing equitonal or equitemporal patterns) may have encouraged imagery for listeners familiar with the excerpt, which could have accounted for the obtained correlation between pitch and temporal conditions. By presenting the pitch and temporal patterns separately, the segments may have produced an artificial dissociation between the two types of information for unfamiliar listeners. Another possible criticism of the first experiment is that the final sounded events contributed heavily to the predictive power of previous descriptions of tonal and rhythmic structure because, within condition, the beginning of each trial was always the same and only the end of each trial differed.

To examine these criticisms, the second experiment used segments that each contained variation in both pitch and temporal information. Segments with both pitch and temporal variation should interfere with an imagery strategy that imposes a pitch or temporal pattern from the original music onto the stimulus. To permit comparisons

with the first experiment, the musical segments were based on the same excerpt. Three types of segments were created, each containing both the original pitch and the original 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 3.

Segments in the *pitch-shift* condition contained the original temporal pattern, but the pitch pattern was shifted in varying amounts (one position to the right in the example of Figure 3), 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, with the pitch pattern being kept constant and the temporal pattern being shifted. Segments in the *combined-shift* condition were created by shifting the pitch and temporal patterns the same amount and in the same direction (essentially choosing a different beginning and ending point). For every segment in the combined-shift condition, there was a segment ending on the same pitch in the pitch-shift condition and a segment ending with the same temporal event in the temporal-shift condition.

Again, listeners who were familiar or unfamiliar with the excerpt rated how good or complete a phrase each segment made. A low correlation between pitch and temporal conditions in this experiment for familiar as well as unfamiliar listeners would support the explanation that the obtained correlation in the first experiment was due to musical imagery for the missing pattern. In this experiment (unlike the previous one), both patterns were always present, to presumably create interference with an image of the original temporal and pitch patterns.

Figure 3. Sample trials for Experiment 2, showing shift of one unit. Shift paradigm demonstrated for melody; accompanying harmony not shown. The three conditions are: (Panel A) pitch-shift condition, (Panel B) temporal-shift condition, and (Panel C) combined-shift condition. Arrows indicate beginning of original pattern of pitch (P1) and temporal (T1) patterns.

Method

Subjects. Sixteen adult listeners were recruited from the Cornell University community. They had had an average of 10.2 years' musical training on an instrument or voice, with a range of 3 to 23 years. None of the listeners had hearing problems, and none had been in the first experiment. Each received \$4/h for participating.

Apparatus and Stimulus materials. Stimuli were generated on the same equipment, and each tone contained the same harmonically complex structure, as in the first experiment. In the pitch-shift condition, all segments began on the first temporal unit (labeled T1 in Figure 3) and ended on the last temporal unit of the original excerpt, but the pitch pattern was shifted and recombined with the temporal pattern. In the temporal-shift condition, all segments began on the first pitch (P1) and ended on the last pitch of the original excerpt, but the temporal pattern was shifted and recombined with the pitch pattern. Segments in the combined-shift condition were created by shifting both the pitch and temporal information the same amount and in the same direction, wrapping around to the first unit (T1,P1), after reaching the last unit. Pitch and temporal shifts were constructed so that they retained the original harmonic information (combination of simultaneously sounded pitches) in all stimuli. All note durations were the same as in the combined condition of the first experiment. As there were 36 events or units, 36 sequences were created for each condition, giving a total of 108 trials.

Procedure. The procedure was the same as in the first experiment. The experiment was run over a 2-day period. Half of the listeners heard the combined-shift condition on Day 1; the other half did so on Day 2. Trials from the pitch and temporal shift conditions, heard on the alternate day, were randomly intermixed.

Results

Unfamiliar listeners. The mean intersubject correlation was .69 ($p < .05$). Simple correlations between ratings from the three conditions are shown in Table 2. The correlations of the combined-shift condition with pitch- and temporal-shift conditions are high. The correlation between the temporal- and pitch-shift conditions, although statistically significant, is much lower. A regression analysis was then performed that predicted the combined-shift data from the pitch-shift and temporal-shift data. The simple additive model provided a good fit ($R = .87$, $p < .001$); both regression coefficients were significant (standardized pitch coefficient = .72, $p < .01$; standardized temporal coefficient = .28, $p < .01$), and the pitch coefficient was larger than the temporal coefficient

[$F(1,34) = 12.92$, $p < .01$]. The simple additive model was again contrasted with several interaction models, for both grouped and individual data; none produced a better fit than the simple additive model.

The predictive power of the previous structural descriptions is also shown in Table 2. Krumhansl and Kessler's (1982) tonal hierarchy for the melody did not correlate significantly with the pitch-shift or the combined-shift ratings, a result that differs from that of the first experiment. The metrical hierarchy did correlate significantly with both the temporal-shift and the combined-shift ratings. The time-span reduction tree correlated significantly with the combined-shift ratings, and the semipartial correlation between combined-shift ratings and time-span reduction, after metrical structure effects were removed, was also significant ($r = -.51$, $p < .01$).

The three sets of predictions were again compared. Krumhansl and Kessler's (1982) ratings of fit for pitches in the melody did not correlate significantly with the metrical structure predictions of Lerdahl and Jackendoff (1983) ($r = .28$, $p > .05$) or with the time-span predictions ($r = -.19$, $p > .20$). The metrical structure and time-span reductions correlated significantly ($r = -.65$, $p < .01$), reflecting the formation of the time-span reduction in part from the metrical structure.

Familiar listeners. The mean intersubject correlation was .41 ($p < .05$) for the familiar listeners. The simple correlations between ratings from the three conditions are shown in Table 2. The correlations of the combined-shift condition with both the pitch-shift and temporal-shift conditions are high. The correlation between pitch- and temporal-shift conditions (.54) is significant, although lower than that obtained in the first experiment (.78). The simple additive model, which predicted the combined-shift data from the pitch- and temporal-shift data, provided a good fit ($R = .79$, $p < .001$). The standardized pitch coefficient was significant (.64, $p < .001$), and the standardized temporal coefficient approached significance (.23, $p < .10$). The pitch coefficient was significantly larger than the temporal coefficient [$F(1,34) = 11.74$, $p < .01$]. The simple additive model compared favorably with several models that assumed an interaction between pitch and temporal factors, for both grouped and individual data.

The tonal and rhythmic predictions are also shown in Table 2. Krumhansl and Kessler's (1982) tonal hierarchy for the melody correlated significantly with the pitch-shift condition, but not the combined-shift condition. Lerdahl and Jackendoff's (1983) metrical structure correlated significantly with both the temporal-shift condition and the combined-shift condition. The time-span reduction tree correlated significantly with the combined-shift ratings, and the semipartial correlation between combined-shift ratings and time-span reduction, after metrical structure effects were removed, was significant as well ($r = -.34$, $p < .05$).

The relatedness of pitch and temporal components was evaluated for the two groups of listeners. If familiar listeners showed correlated pitch and temporal ratings in

Table 2
Simple Correlations in Experiment 2
as a Function of Listener Group

	Conditions			Predictions		
	Pitch	Temporal	Combined	Tonal	Metrical	Time-span
Unfamiliar Listeners						
Pitch			.83†	.20		
Temporal			.56†		.64†	
Combined				.30*	.38†	-.72†
Familiar Listeners						
Pitch			.76†	.35†		
Temporal			.57†		.52†	
Combined				.22	.39†	-.57†

Note—All $df = 34$. * $p < .10$. † $p < .05$.

the previous experiment due to imagery, then their simple correlations should be reduced in this experiment and be similar to those of the unfamiliar listeners. This prediction was supported; there were no group differences in the listeners' simple correlations of pitch and temporal ratings [$t(14) = 0.36, p > .70$], and the mean individual correlations were low in both groups (average familiar $r = .16$; unfamiliar $r = .18$).

Comparisons were made between the first and second experiments. A correlation was performed on the average ratings between stimulus conditions. The correlations, shown in Table 3, are significant for both groups of listeners, indicating substantial agreement for ratings on one type of information both in the absence (Experiment 1) and in the presence (Experiment 2) of the other type of information.

Discussion

This experiment verified with a different methodology the previous finding that phrase judgments based on pitch and temporal information are sufficient to predict phrase judgments for the combined information in a Mozart sonata theme. As in the first experiment, a simple additive model predicted the combined-shift ratings from pitch-shift and temporal-shift ratings. This result demonstrates additive contributions of pitch and temporal factors to phrase judgments of the original excerpt, for both familiar and unfamiliar listeners. There was significant agreement between corresponding conditions in the two experiments, supporting additivity of the two patterns, both in the presence and in the absence of the other pattern.

For unfamiliar listeners, the pitch and temporal conditions were weakly correlated, and for familiar listeners, they were less correlated in this experiment than in the previous experiment. There were also no group differences in the individual correlations between pitch and temporal ratings in this experiment, in contrast to the group differences in the first experiment. These findings, along with the uncorrelated predictions for tonal and metrical hierarchies, suggest that the two perceptual hierarchies of rhythm and tonality do not coincide in the Mozart sonata theme. The presence of both types of information in the second experiment may have prevented the imagery that the familiar listeners had experienced in the first experiment.

Table 3
Simple Correlations Between Conditions Across Experimental Paradigms as a Function of Listener Group

Experiments	Condition		
	Pitch	Temporal	Combined
Experiments 1 and 2			
Unfamiliar	.77*	.70*	.86*
Familiar	.79*	.76*	.92*
Experiments 3 and 4			
Unfamiliar	.91*	.64*	.84*
Familiar	.79*	.49*	.94*

Note—All $df = 15$. * $p < .05$.

Predictions of tonal and rhythmic hierarchies explained a significant amount, but not all, of the variance in phrase judgments. The ratings of tonal hierarchies from Krumhansl and Kessler (1982) did not correlate well with the results in this experiment, possibly because the tonal profiles correspond only to the melody and, in this paradigm, harmonic (multivoiced) information was present in every segment. The metrical structure from Lerdahl and Jackendoff (1983) was supported by the experimental results of all listeners. Although intended only for the original combination of pitch and temporal patterns and not for the shifted segments, the time-span predictions were found to correspond to the combined-shift ratings.

In summary, no emergent perceptual qualities appeared for musical phrases containing both pitch and temporal patterns that were not found in the ratings of the separate components. Furthermore, there were no differences in the predictive power of the additive model associated with musical training or familiarity, indicating a large range of listeners for whom temporal and pitch hierarchies operate additively in phrase determination.

The mechanically regular stimuli used in these experiments may induce a perceptual relationship between pitch and temporal information that is different from that induced by live musical performances. For example, a live piano performance includes temporal deviations from the musical score, such as slowing down at cadences or structural endings. Pitch or harmonic changes in the score, as well as temporal changes, also indicate cadences and signify endings of musical phrases. This example of multiple sources of cadential information, from both performed temporal deviations and notated temporal and pitch information, suggests that performance timing may induce a coupling of pitch and temporal information.

Although frequency and timbre are largely fixed on keyboard instruments, there are several ways pianists can manipulate durations and intensity. Early studies of timing in piano performance (Henderson, 1937) demonstrated that relative durations of notes were correlated with traditional accent placement (the first beat of each measure). Piano recordings indicated that notes preceding accented positions were lengthened, serving to delay their entrance. More recently, Sloboda (1983) studied expressive variations in piano performance and found that rubato (tempo changes), legato and staccato (note offset-to-note onset durations), and intensity patterns were used to accent important note events. Moreover, the use of these methods changed as the metrical accent placement in the musical notation was changed.

Studies of performances by singers (Seashore, 1937), flutists (Bengtsson & Gabrielsson, 1983), and drummers (Gabrielsson, 1974) suggest that pianists' methods of performance timing are not unique, and may reflect general perceptual mechanisms for parsing acoustic material into units such as phrases. Specifically, lengthening of durations at phrase boundaries may facilitate perception of a musical composition's hierarchical structure. Analysis of note durations in singers' performances indicated that

pauses between phrases were found to be four times longer than pauses within phrases (Seashore, 1937). Todd (1985) developed a model to generate interphrasal durations by an amount proportional to the hierarchical level or depth of phrase embedding; the model compared favorably at the phrase level with three piano performances. Restle (1972) investigated the role of pausing in phrasing of light patterns. Learning of sequential light patterns was facilitated most when long pauses corresponded to major divisions and short pauses to minor divisions in the hierarchical tree description. He argued that phrasing in speech and music, as well as in light patterns, serves to facilitate learning of hierarchical organization.

Performance timing deviations are found within as well as between phrases, in both monophonic (Gabrielsson, 1974) and polyphonic (Palmer, 1986) piano performances. The performances were characterized by note durations that varied systematically from the strict mechanical regularity of the notated score. Rhythms notated as integer ratios were generally made larger, such that the longer durations were even longer and the short durations shorter (Gabrielsson, 1974). Also, durations of events at structural endings were lengthened relative to durations within the structure. Moreover, the duration deviations in performances of the same sonata excerpt used in the present study were larger than 10% (Palmer, 1986), the just noticeable perceptual difference for tones and lights of the same durations (Woodrow, 1951).

Another method of performance timing, related to harmonic organization, is an asynchrony of events notated as simultaneous, or chord asynchrony. An asynchrony between chamber ensemble players was reported by Rasch (1979), who found a small lead (less than 10 msec) of the instrument playing the melody. The leading instrument may serve to coordinate timing between ensemble players, or a leading melody may increase the perceptual salience of that musical line. Chord asynchronies were also found in piano performances (Palmer, 1986) for which no coordinating leader is required. The asynchronies were greatly reduced when pianists were asked to play unmusically, indicating that performers had control of the asynchronies. There is perceptual evidence that supports the salient role of a leading melody; Vernon (1937) found that experienced musicians could perceive temporal asynchronizations as small as 10 msec between the onsets of 2 notes on a piano. Accurate perceptual ordering of two successive events occurs for asynchronies of 10 to 20 msec, for both tones and lights (Hirsh, 1959; Hirsh & Sherrick, 1961).

Because these methods of performance timing serve to highlight events of structural importance, accentuate phrasing, and articulate melodic lines, it is conceivable that they would also affect phrase judgments in tasks such as those used in the present studies. In particular, the timing deviations coupled to the tonal information suggest that pitch and temporal variables may interactively affect the perceptual organization of a live performance. Therefore, we repeated the first two experiments, basing the

stimulus materials' durations on the timing from a live performance. To acquire a precise record of the performance, a concert pianist was asked to perform the Mozart excerpt on a synthesizer monitored by a computer. The temporal information associated with the live performance (note onset and offset times) was used to generate the musical stimuli in the third and fourth experiments. If performance timing deviations are closely linked with pitch information, as the studies described above suggest, then the additive relationship of pitch and temporal information in phrase judgments may be disrupted by performance timing. Alternatively, if temporal and pitch information are truly processed independently, then performance timing should not alter their relationship from that obtained in the first two experiments.

EXPERIMENT 3: ADDING PATTERNS WITH PERFORMANCE TIMING

The third experiment investigated whether or not judgments of musical phrases could be predicted from the separated pitch and temporal information when performance timing was included. Segments were created from the same piano sonata theme (Figure 1). With the exception of the note durations, the three types of segments produced were the same as in the first experiment (shown in Figure 2). 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 duration pattern from the live performance, but all tones had the same pitch (equitonal); and segments in the *combined condition* retained both the original pitch pattern and the performed duration pattern. Listeners either unfamiliar or familiar with the excerpt rated how good or complete a phrase each segment made.

If perceived goodness of a phrase is a function of pitch and temporal information, regardless of the performance deviations, then ratings in the combined condition should be predicted accurately from a combination of the ratings from the pitch and temporal conditions. The degree of association between pitch and temporal factors may also be affected by performed timing.

Method

Subjects. Sixteen adult listeners from the Cornell University community were recruited. The average number of years of formal training on an instrument or voice was 8.6; the range was from 2 to 27 years. None had hearing problems, and none had been in any of the other experiments. Each was paid \$4/h for participating.

Apparatus and Stimulus materials. A live performance of the excerpt was recorded on a Roland Juno-106 electronic synthesizer, which used digital-waveform synthesis combined with analog filters, and constant amplitude output for all events (fixed loudness level). The synthesizer was monitored by an IBM-XT personal computer. A Roland MPU-401 MIDI interface interpreted and timed all note events from the synthesizer with a timing resolution of 2 msec and precision (standard deviation) of 0.6% for the range of durations recorded. The synthesizer's output was sent to a Superscope A-240 amplifier and a Dynaco speaker placed directly in front of the syn-

thesizer. The synthesizer's timbre was a 33% pulse waveform with a high-level (80%) low-pass filter. The rise time of the amplitude envelope was 10 msec, with a negative exponential decay (fitted by a linear ramp in the four experiments) to zero amplitude over the duration minus the release time. The release time (time from pianist's release of a key to zero amplitude) was 80 msec.

The note durations from the pianist's performance are displayed in Figure 4 in terms of percent deviation from a regular (mechanical) performance. The mean tempo was 217 msec per sixteenth note, or 46 beats per minute (the same tempo used in all of the experiments). The pianist held a bachelor's degree in music from the Curtis Institute. She had had 8 years' experience concertizing in America and Europe, performing both solo and chamber works from the Western classical and contemporary repertoire. The instruments with which she was familiar included: modern piano, harpsichord, and flute. The pianist was asked to perform the excerpt, with which she was familiar, as often as she wished. After hearing each recording, played back by the computer, she chose the one she felt most accurately represented her interpretation.

There are several patterns in the live performance which suggest a correspondence between pitch information and performance timing. First, the largest deviations in Figure 4 (representing passages at which the tempo slowed down) are at the cadences in measures 4 and 8. Cadences are also typically indicated by pitch or harmonic movement toward more stable chords, such as the V chord at the last event in measure 4 and the I chord at the last event in measure 8. Thus, the performance timing and pitch information are coupled at the cadences. Second, the shortened durations (largest negative deviations) correspond to passing tones, or pitches of least structural importance. Finally, the relative asynchrony between tones notated as simultaneous (chords) is significantly larger than zero. The melody (in this excerpt, the highest pitch in each chord) preceded the other voices in all three-note chords by an average of 15 msec, significantly larger than zero [$t(25) = 7.8, p < .001$].

All stimuli were generated as in the first experiment, except that the tones' durations and relative onsets and offsets were set to match those measured from the live performance, with one exception. Offsets overlapping with following note onsets were set equal to the onset of the next tone (due to computer memory limitations). This

did not result in any major acoustic differences, because the final amplitude was quite diminished by the onset of the next tone.

Procedure. The procedure was identical to that of the first experiment.

Results

Unfamiliar listeners. The mean intersubject correlation was .36 ($p < .10$) for unfamiliar listeners. Simple correlations between average ratings from the three conditions are shown in Table 4. The correlation between pitch and combined conditions was significant, as was the correlation between temporal and combined conditions. The correlation between pitch and temporal conditions was, however, not significant. A regression analysis that predicted the combined-condition data from the pitch- and temporal-condition data provided an excellent fit ($R = .92, p < .001$). Both regression coefficients were significant (standardized pitch coefficient = .59, $p < .01$; standardized temporal coefficient = .49, $p < .01$), and the coefficients did not differ significantly [$F(1,15) = 0.02, p > .80$]. The additive combination of pitch- and temporal-condition judgments in predicting combined-condition judgments compared favorably with several interaction models, on both a group and individual basis.

The fits of the tonal and rhythmic predictions are also shown in Table 4. Krumhansl and Kessler's (1982) tonal hierarchy data correlated significantly with both the pitch-condition and the combined-condition ratings. The metrical structure of Lerdahl and Jackendoff (1983) correlated significantly with the temporal-condition ratings, and the correlation with the combined-condition ratings approached significance. The time-span predictions correlated significantly with the combined-condition ratings, and the semipartial correlation of time-span predictions

Figure 4. Duration deviations based on a live performance of the excerpt. Horizontal line indicates a mechanically regular performance.

and combined-condition ratings, after metrical-structure effects were partialled out, was also significant ($r = -.52, p < .05$).

Familiar listeners. The mean intersubject correlation was .53 ($p < .05$) for familiar listeners. Simple correlations between ratings from the three conditions are shown in Table 4. The correlation between pitch and combined conditions was significant, as was the correlation between temporal and combined conditions. The correlation between pitch and temporal conditions was also significant, indicating strong agreement between those ratings. The simple additive model that predicted the combined-condition data from the pitch- and temporal-condition data again provided an excellent fit ($R = .96, p < .001$). Both regression coefficients were significant (standardized pitch coefficient = .69, $p < .001$; standardized temporal coefficient = .31, $p < .05$), and they did not differ significantly [$F(1,15) = .08, p > .70$]. The additive combination of pitch- and temporal-condition judgments in predicting combined-condition judgments was compared with several interaction models, for both grouped and individual data; no other model provided a better fit.

The tonal and rhythmic predictions, shown in Table 4, again provided a good fit. Krumhansl and Kessler's (1982) key-context data correlated significantly with both the pitch- and the combined-condition ratings. Predictions from the metrical structure of Lerdahl and Jackendoff (1983) correlated significantly with temporal-condition ratings, and the correlation approached significance with the combined-condition ratings. Finally, the time-span reduction correlated with the combined-condition ratings, and the semipartial correlation of combined-condition ratings and time-span reduction, after metrical structure effects were partialled out, was also significant ($r = -.52, p < .05$).

Differences between familiar and unfamiliar listeners in the correlations of temporal and pitch ratings were evaluated. The simple correlation was higher on average for familiar ($r = .66$) than for unfamiliar ($r = .33$) listeners [$t(14) = 2.46, p < .05$]. Subject reports of imagery, along with parallel findings in Experiment 1 (using

Table 4
Simple Correlations in Experiment 3
as a Function of Listener Group

	Conditions			Predictions		
	Pitch	Temporal	Combined	Tonal	Metrical	Time-span
Unfamiliar Listeners						
Pitch		.45	.81†	.52†		
Temporal			.75†		.61†	
Combined				.49†	.47*	-.74†
Familiar Listeners						
Pitch		.83†	.94†	.54†		
Temporal			.88†		.67†	
Combined				.59†	.56*	-.77†

Note—All $df = 15$. * $p < .10$. † $p < .05$.

Table 5
Simple Correlations Between Experimental Conditions
With and Without Performance Timing

Experiments	Condition		
	Pitch	Temporal	Combined
Experiments 1 and 3			
Unfamiliar	.89*	.75*	.80*
Familiar	.90*	.95*	.90*
Experiments 2 and 4			
Unfamiliar	.82*	.59*	.79*
Familiar	.79*	.38*	.80*

Note—Experiments 1 and 3, $df = 15$; Experiments 2 and 4, $df = 34$. * $p < .05$.

the same methodology without performance timing), suggest that the group difference is due to imagery rather than performance timing.

Effects of performance timing on the phrase ratings were evaluated by comparing Experiments 1 (without performance deviations) and 3 (with performance deviations). The simple correlations between corresponding conditions across the two experiments are shown in Table 5. All conditions correlated significantly, indicating a high degree of consistency between experiments whether or not performance timing was absent (Experiment 1) or present (Experiment 3).

Discussion

This experiment demonstrated that phrase judgments of a harmonic excerpt can be described in terms of ratings based on the separate pitch and temporal patterns, and this relation is unaffected by performance timing. In particular, an additive model (without an interaction term) was sufficient, suggesting, along with Experiment 1, that pitch and temporal factors are additive in phrase judgments for both mechanically regular (as notated) and deviating (as performed) temporal patterns. Ratings in the pitch and temporal conditions did not correlate for unfamiliar listeners. This result indicates that perceptually strong tonal events are not paired with metrically strong beats in the theme for unfamiliar listeners.

As in the first experiment, the association between temporal and pitch ratings was stronger for familiar than for unfamiliar listeners. The fact that this difference was reduced in the shift paradigm of Experiment 2, along with listeners' self-reports, suggests that imagery, rather than performance timing, influenced familiar listeners' phrase judgments.

Finally, the influence of previously described tonal and rhythmic structures was again evident. All sets of predictions (Krumhansl & Kessler's, 1982, tonal hierarchy, Lerdahl & Jackendoff's, 1983, metrical and time-span structures) correlated well with ratings, for both groups of listeners. These accurate predictions indicate the importance of events at phrase endings, and the predictions are robust for both performed and mechanically regular timing.

EXPERIMENT 4: SHIFTING PATTERNS WITH PERFORMANCE TIMING

A fourth experiment, using the shift methodology of the second experiment, was designed to further test the contributions of pitch and temporal factors in the presence of performance timing. In all three conditions of this experiment (constructed as in Experiment 2), the temporal pattern matched that of the live performance (also used in Experiment 3). Again, both listeners familiar and listeners unfamiliar with the excerpt rated how good or complete a phrase each segment made.

If the simple additive model were to predict phrase judgments from combinations of pitch and temporal information in spite of the performance deviations, then combined-shift judgments should be fit well by a linear combination of the corresponding pitch- and temporal-shift judgments. If the correlation between the pitch- and temporal-shift conditions was found to be low for familiar as well as unfamiliar listeners, then this should provide converging evidence that imagery, rather than performance timing, contributed to the correlation previously obtained.

Method

Subjects. Sixteen adult listeners were recruited from the Cornell University community. They had had an average of 9.3 years' formal instruction on an instrument or voice, with a range of 4 to 22 years. None had hearing problems or had been in any of the other experiments. Each was paid \$4/h for participating.

Apparatus, Stimulus materials, and Procedure. Stimuli were generated using the same methods and equipment as in the second experiment. The stimuli used in the three conditions were identical to those of Experiment 2 (shown in Figure 3), with the addition of the performance timing used in Experiment 3. The procedure was the same as in the second experiment.

Results

Unfamiliar listeners. The mean intersubject correlation was .43 ($p < .05$) for unfamiliar listeners. Simple correlations between ratings from the three conditions are shown in Table 6. The correlations of the combined-shift condition with both pitch- and temporal-shift conditions are high. The correlation between pitch- and temporal-shift conditions is lower, although significant. The sim-

ple additive model predicting the combined-shift data from the pitch- and temporal-shift data provided a good fit ($R = .82, p < .001$); both regression coefficients were significant (standardized pitch coefficient = .49, $p < .001$; standardized temporal coefficient = .47, $p < .001$), and they did not differ significantly [$F(1,34) = .41, p > .50$]. The simple additive model was again contrasted with several models that assumed an interaction between pitch and temporal factors, both on individual and group data; none of the other models improved the fit.

The fit of the predictions is also given in Table 6. Krumhansl and Kessler's (1982) tonal hierarchy correlated significantly with the combined-shift condition, but the correlation with the pitch-shift condition was not significant. Lerdahl and Jackendoff's (1983) metrical-structure predictions correlated significantly with both the temporal- and combined-shift conditions. The time-span reduction correlated significantly with the combined-shift ratings; however, the semipartial correlation between combined-shift ratings and time-span reduction, after metrical structure effects were removed, was not significant ($r = -.21, p > .20$), indicating that the time-span reduction did not add, in this case, beyond the metrical structure.

Familiar listeners. The mean intersubject correlation was .37 ($p < .05$) for familiar listeners. Simple correlations between ratings from the three conditions are shown in Table 6. The correlations of the combined-shift condition with both pitch- and temporal-shift conditions were significant. The correlation between pitch- and temporal-shift conditions was also significant ($r = .47$) but lower than that obtained in Experiment 3 ($r = .83$), suggesting that the imagery allowed by the previous paradigm, rather than the performance durations, contributed in Experiment 3. The simple additive model that predicted the combined-shift data from the pitch- and temporal-shift data again provided a good fit ($R = .74, p < .001$). Both regression coefficients were significant (standardized pitch coefficient = .50, $p < .001$; standardized temporal coefficient = .36, $p < .01$), and they did not differ significantly [$F(1,34) = 1.95, p > .15$]. The simple additive model was again contrasted with several models that assumed an interaction between pitch and temporal factors for both group and individual data; none provided a better fit than the additive model.

The fits of the tonal and rhythmic predictions are also shown in Table 6. Krumhansl and Kessler's (1982) tonal hierarchy did not correlate significantly with the combined-shift condition and approached significance with the pitch-shift condition. Lerdahl and Jackendoff's (1983) metrical structure did not correlate significantly with either the temporal- or the combined-shift conditions. The time-span reduction correlated significantly with the combined-shift ratings, and the semipartial correlation between combined-shift ratings and time-span reduction, after metrical structure effects were removed, was significant ($r = -.41, p < .05$).

Differences between familiar and unfamiliar listeners were evaluated by examining the individual correlations of pitch- and temporal-shift ratings for the two groups.

Table 6
Simple Correlations in Experiment 4
as a Function of Listener Group

	Conditions			Predictions		
	Pitch	Temporal	Combined	Tonal	Metrical	Time-span
Unfamiliar Listeners						
Pitch		.45†	.70†	.30*		
Temporal			.69†		.34†	
Combined				.38†	.46†	-.49†
Familiar Listeners						
Pitch		.47†	.67†	.31*		
Temporal			.59†		-.12	
Combined				.21	.20	-.53†

Note—All $df = 34$. * $p < .10$. † $p < .05$.

Only one of the 16 listeners' simple correlations between temporal- and pitch-shift ratings was significant. Thus, there was little association between temporal and pitch components for either familiar (average $r = .27$) or unfamiliar ($r = .13$) listeners.

Comparisons between experiments are given in Tables 3 and 5. All correlations between corresponding conditions across the third and fourth experiments were significant for both groups of listeners (Table 3), indicating high consistency both in the absence (Experiment 3) and in the presence (Experiment 4) of one of the patterns of information. In addition, the effects of the live performance were evaluated by comparing Experiments 2 (without performance) and 4 (with performance). The simple correlations between corresponding conditions for these experiments are given in Table 5. All correlations were significant, again indicating strong consistency in ratings for segments with or without performance deviations.

Discussion

This experiment verified that phrase judgments based on pitch and temporal information are sufficient to predict judgments based on a live performance of the sonata theme. As in the third experiment, the simple additive model, predicting combined-shift ratings from pitch- and temporal-shift ratings, was sufficient; models with interaction terms provided no better fits. The significant agreement between ratings in Experiments 2 (without performance durations) and 4 (with performance durations) indicates that the contributions of pitch and temporal information to phrase judgments are unaffected by performance timing deviations from the notated score.

The pitch- and temporal-shift judgments did not correlate for unfamiliar listeners, and the correlation was reduced from Experiment 3 for familiar listeners. There were also no group differences for the individual correlations of temporal and pitch judgments. These results were also found in Experiment 2, which was based on the same imagery-discouraging methodology. Thus, the relatedness of temporal and pitch components appears to be influenced by conditions favorable to imagery rather than by performance timing variations.

The tonal and rhythmic hierarchy predictions explained a significant amount of the variance in unfamiliar listeners' judgments. The predictions may not have fit the familiar listeners' ratings as well because performance timing mismatches to a familiar piece may be more disruptive and may result in more variability in ratings. This interpretation is supported by the lower intersubject correlations and lower multiple regression coefficients found for familiar than for unfamiliar listeners.

GENERAL DISCUSSION

The present findings indicate that phrase judgments in a Mozart sonata theme are based on two additive types of structure: pitch and temporal information. This con-

clusion was supported by experiments that used two different paradigms, each of which preserved one pattern of information while altering the other, and performance timing as well as mechanically regular timing. Furthermore, the generality of the additive model across levels of musical training and familiarity with the excerpt indicates a large range of listeners for whom pitch and temporal hierarchies operate additively in phrase determination.

The relationship between pitch and temporal components differed with familiarity with the music, but less so when the experimental design discouraged imagery. When one component was presented without the other, listeners familiar with the excerpt reported imagery for the missing component and tended to have higher correlations between pitch and temporal ratings than did listeners unfamiliar with the excerpt. When both components were present, presumably preventing imagery, there were no reports of imagery and familiar and unfamiliar listeners showed the same (low) degree of association between the components. The correlation was reduced, both on average and on an individual basis. This suggests that imagery contributed to the correlation, and that once imagery is discouraged, the components may be processed independently in phrase judgments. Grouping principles that contribute to phrase judgments may be general auditory processes that operate separately on pitch and temporal information in music, as well as in the perceptual organization of other stimulus dimensions. These findings also support the more general theory that listeners' memory representations for familiar music contain both temporal and pitch structure and that hearing only one structure is sufficient for recognition.

Previous descriptions of tonal and rhythmic structures explained 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) generally correlated highly with the ratings as a function of the last-sounded pitch in the melody. This perceived structure, previously measured with abstract key-defining contexts, also operates in the more complex harmonic conditions of the sonata theme. The metrical and time-span predictions of Lerdahl and Jackendoff (1983) also agreed with the phrase ratings, suggesting that rules of time-span reduction may be additive, rather than interactive, for music of the classical sonata form.

The performance timing applied in the last two experiments did not alter the relationship between pitch and temporal components in predicting phrase judgments. In fact, temporal and pitch information may require separate perceptual processes in part because of constraints on the production of timed intervals. If perception of temporal and pitch components were greatly affected by performed timing variations, then random variability in production of timed intervals (from sources such as neural limits on motor precision) would cause a perceptual interaction when none was intended. Even when attempting to per-

form with mechanical regularity (without any timing variations), musicians display some timing variability. Additionally, if performed timing variations were sufficient to cause a perceptual interaction between pitch and temporal factors, then different performances of the same piece would provide a very different perception. However, recognition of a melody's identity is quite robust across different performances.

The additivity of pitch and temporal factors in musical phrase judgments of the classical period, as well as Bach's period (Palmer & Krumhansl, 1987), suggest that the relationship between pitch and temporal structures does not necessarily change for music that places a greater emphasis on vertical (harmonic) organization. Harmonic music that contained multiple pitch lines did not encourage listeners' organized percepts to be based on an interaction of tonal and rhythmic components. In fact, the fit of Krumhansl and Kessler's (1982) tonal hierarchy for the melody suggests that chordal information may be perceived in terms of the individual components as well as their combination; there was no failure to attend to individual parts, which a perceptual interaction might predict. These findings suggest that even in this classical excerpt with vertical organization, horizontal organization is perceived, and perception of musical phrases in this style is influenced by individual horizontal structures as well as their vertical intersections.

REFERENCES

- BENGTSSON, I., & GABRIELSSON, A. (1983). Analysis and synthesis of musical rhythm. In J. Sundberg (Ed.), *Studies of music performance, publications issued by the Royal Swedish Academy of Music*, 39, 27-60.
- CLARKE, E. F. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross, & R. West (Eds.), *Musical structure and cognition* (pp. 209-236). London: Academic Press.
- DEUTSCH, D. (1980). The processing of structured and unstructured tonal sequences. *Perception & Psychophysics*, 28, 381-389.
- GABRIELSSON, A. (1973). Similarity ratings and dimension analyses of auditory rhythm patterns. I. *Scandinavian Journal of Psychology*, 14, 138-160.
- GABRIELSSON, A. (1974). Performance of rhythm patterns. *Scandinavian Journal of Psychology*, 15, 63-72.
- HENDERSON, M. T. (1937). Rhythmic organization in artistic piano performance. In C. E. Seashore (Ed.), *Objective analysis of musical performance, University of Iowa studies in the psychology of music IV* (pp. 281-305). Iowa City: University of Iowa Press.
- HIRSH, I. J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America*, 31, 759-767.
- HIRSH, I. J., & SHERRICK, C. E. (1961). Perceived order in different sense modalities. *Journal of Experimental Psychology*, 62, 423-432.
- 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., & KESSLER, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, 89, 334-368.
- LERDAHL, F., & JACKENDOFF, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.
- MONAHAN, C. B., & CARTERETTE, E. C. (1985). Pitch and duration as determinants of musical space. *Music Perception*, 3, 1-32.
- PALMER, C. (1986). Methods of articulation in piano performance. *Journal of the Acoustical Society of America*, 79, S75.
- PALMER, C., & KRUMHANSL, C. L. (1987). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception & Performance*, 13, 116-126.
- RASCH, R. A. (1979). Synchronization in performed ensemble music. *Acustica*, 43, 121-131.
- RESTLE, F. (1972). Serial patterns: The role of phrasing. *Journal of Experimental Psychology*, 92, 385-390.
- SEASHORE, H. G. (1937). An objective analysis of artistic singing. In C. E. Seashore (Ed.), *Objective analysis of musical performance, University of Iowa studies in the psychology of music IV* (pp. 12-157). Iowa City: University of Iowa Press.
- SHAFFER, L. H., CLARKE, E. F., & TODD, N. P. (1985). Metre and rhythm in piano playing. *Cognition*, 20, 61-77.
- SLOBODA, J. A. (1983). The communication of musical metre in piano performance. *Quarterly Journal of Experimental Psychology*, 35A, 377-396.
- TODD, N. (1985). A model of expressive timing in tonal music. *Music Perception*, 3, 33-58.
- VERNON, L. N. (1937). Synchronization of chords in artistic piano music. In C. E. Seashore (Ed.), *Objective analysis of musical performance, University of Iowa studies in the psychology of music IV* (pp. 306-345). Iowa City: University of Iowa Press.
- WHITE, B. W. (1960). Recognition of distorted melodies. *American Journal of Psychology*, 73, 100-107.
- WOODROW, H. (1951). The perception of time. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley.