Cue Trading in the Perception of Rhythmic Structure
Author(s): Hasan Gürkan Tekman
Source: *Music Perception: An Interdisciplinary Journal*, Vol. 13, No. 1 (Fall, 1995), pp. 17-38
Published by: University of California Press
Stable URL: http://www.jstor.org/stable/40285683

Music Perception
Fall 1995, Vol. 13, No. 1, 17–38

Cue Trading in the Perception of Rhythmic Structure

HASAN GÜRKAN TEKMAN
Middle East Technical University, Turkey

The effects of variations in three pairs of variables on identification and discrimination of rhythmic patterns in pure-tone sequences were investigated. It was found that adding a timing difference to two sequences that differed in terms of the pattern of tone intensities improved discrimination if variations of the two variables converged on the same rhythmic pattern but did not help if the variation in timing undermined the rhythmic pattern created by the intensity variation. Adding a difference in the intensities of tones to two sequences that differed in terms of the pitch structure produced a similar pattern of results. The effect of adding a difference in pitch structure to two sequences that differed in terms of timing was not reliably related to the way the differences were combined. The results showed that, at least for some variables, predicting discrimination performance from probabilities of detecting a rhythmic pattern is possible. The relationship of the results to cue trading in speech perception and research possibilities with similar methods are discussed.

MANY activities related to sound and movement make use of time in a very structured manner. In music, this organization is described by terms such as beats, accents, meter, and tempo. These concepts refer to abstractions made by listeners on the basis of physical stimuli. The distinction between phenomenal accents and metrical accents that was made by Lerdahl and Jackendoff (1983) in their *Generative Theory of Tonal Music* nicely illustrates this point. Phenomenal accents are events that give emphasis to moments in the musical flow. Metrical accents, on the other hand, are beats that are relatively strong in their metrical contexts. Thus, phenomenal accents are physical events from which metrical accents, which are abstractions, are derived.

Theories of temporal structure in music generally involve the assumption that temporal structure is determined by a variety of cues in alternation and in combination (Jones, 1976; Jones & Boltz, 1989; Lerdahl &

Requests for reprints may be addressed to Hasan Gürkan Tekman, Department of Psychology, Middle East Technical University, ODTU 06531, Ankara, Turkey. (e-mail: tekman@rorqual.cc.metu.edu.tr)

Jackendoff, 1983; Tenney, 1964/1988; Tenney & Polansky, 1980). Lerdahl and Jackendoff, for example, consider attack points of notes, notes that are louder or longer than their neighbors, notes that initiate groups united by a slur, the same dynamic marking, or the same manner of articulation as determiners of temporal groups. Tenney, whose theory of perceptual grouping in music that is not constrained by the use of the traditional metric structure of Western music (Tenney, 1964/1988; Tenney & Polansky, 1980), and Jones (1976), who used the concept of rhythmic structure in a theory of attention that emphasized the temporal deployment of attention, also assume that temporal structure is determined by larger separation in time or larger changes in other parameters of sound such as intensity, pitch, and timbre.

One important point about these theories of perceived temporal structure (Jones, 1976; Lerdahl & Jackendoff, 1983; Tenney, 1964/1988) is that different types of events can substitute for each other in giving structure to time or combine to mark higher order boundaries. Indeed, it was found that when two accenting effects converge at the same point in time, their combined effect is larger than their individual effects (Jones, 1992).

The main goal of the research presented in this article was to find out whether differences in one dimension of sound could go undetected as a result of compensating changes in other dimensions. Interactions of dimensions of sound with each other as they relate to rhythmic structure have been observed in such phenomena as departures from written music in performance, perception of such expressive deviations by listeners, and perceptual interactions in nonmusical stimuli. As the research reviewed below demonstrates, people have the experience of and preference for hearing expressive variations on multiple dimensions of sound that communicate temporal structure in music, and they also have the illusion of hearing multiple variations when physically only one is present.

Expressive Variations in Performance

Combination of multiple cues to communicate temporal structure has been observed in detailed analyses of music performance. Performers depart from strict adherence to written music in ways that communicate the temporal structure of the piece more effectively. Gabrielsson (1974) asked musicians to reproduce written rhythms on a piano or drums. He found that his subjects departed from the notated rhythm in order to accentuate the metrical accent points. This was accomplished by a combination of effects: The accented notes were played louder, the beat period corresponding to the note (the time period starting with the onset of the note and ending with the onset of the note corresponding to the next beat) was lengthened, and the pause that followed the accented note was shortened. Sloboda (1983) observed similar variations from notated music with pianists sight-reading short passages. Furthermore, he observed that these variations facilitated listeners' identification of the metrical structure of the pieces.

In an attempt to create computer-controlled performances of music that would sound "natural" to listeners, a group of researchers found it necessary to implement variations in durations, intensities, and fine tuning of pitches that were not indicated in the written score (Friberg, 1991; Friberg, Frydén, Bodin, & Sundberg, 1991; Sundberg, Friberg, & Frydén, 1991). Particular patterns of change in one dimension of sound created expectancies for particular variations in other dimensions. Friberg and his associates created a computer program that added interpretive variations, taking its cues from the pitch and timing structure of the notated form of music such that the product was preferred to deadpan performances by listeners.

Perceptual Interactions

Empirical investigations found evidence for the possibility of counteracting rhythmic grouping effects created by a change in one dimension of sound by a change in another dimension. In one early series of experiments on this subject, Woodrow (1909) observed the changes in the grouping structure of a monotone sequence in response to changes in the physical intensity or duration of alternating tones. He found that a sequence of alternating high- and low-intensity sounds was perceived as groups of two sounds that started with the high-intensity sound and ended with the low-intensity sound. This effect of intensity differences could be eliminated and even reversed by making the interonset interval (the interval that separates the onset of one tone from the onset of the following tone) before the high-intensity sound shorter than the interonset interval before the low-intensity sound.

Woodrow (1909) also conducted a similar experiment in which long and short tones alternated in a monotone sequence. With the longer sounds considerably longer than the shorter sounds, when the onsets of the two sounds were equally spaced, subjects reported hearing groups of two, starting with the longer tone and ending with the shorter one. Grouping could be eliminated or reversed by making the interonset interval preceding a short tone longer than the interonset interval preceding a long tone.

As another demonstration of how the variations of one dimension can affect the perception of another dimension, Povel and Okkerman (1981) looked at the effect of temporal patterning of equitone sequences on the perception of accents. In this case, all the tones in the sequences were identical but the durations of the intertone intervals (the intervals that separate the offset of one tone from the onset of the following tone) were varied. A pattern of simple alternation of long and short intertone intervals led the tones that followed the longer intervals to be perceived as accented when the difference between the long and short intervals was small relative to the durations of the intervals. When the relative difference was large, the tones that preceded the longer intervals were perceived as accented. When subjects adjusted the intensities of the tones such that they sounded equally loud, the tones that were perceived to be accented had lower intensities than the other tones. Thus, the perceptual effect of a change in one dimension of sound can be counteracted by a change in another dimension.

Another project explored the possibility of creating a sense of accent that could be substituted for an intensity accent with the use of changes in pitch. Thomassen (1982) had subjects listen to monotone sequences in which every third tone had a higher intensity than the tones in between. At one point in this sequence, this intensity accent was omitted and changes in the frequency of tones were introduced. The sequence was perceived to be regularly accented with the highest probability if the note that represented the largest pitch jump was the tone that was expected to have higher intensity.

My own research (Tekman, 1993) has shown that changes in intensities and relative pitches of evenly spaced tones in pure tone sequences systematically affected the perception of relative timing and intensities, respectively, of those tones. Specifically, intertone intervals that preceded higher intensity tones were perceived to be longer than intertone intervals that preceded lower intensity tones in the same sequence, and tones that terminated larger changes in pitch were heard as having higher intensities than tones that terminated smaller changes in pitch in the same sequence. However, in the same series of experiments, timing of tones was not found to change the perception of pitch in a systematic way.

These pieces of empirical evidence show that, at least for some dimensions of sound, manipulation of one dimension of sound can create changes in perception of other dimensions that are consistent with the accenting and grouping structure that were created by the physical change. The explanation of these phenomena that was favored by the earlier researchers was in line with a Gestalt approach to rhythmical grouping. Fraisse (1956, 1978) and Woodrow (1909, 1951) thought of perception of rhythmic patterns as assimilating groups of sounds into a set of existing mental structures. Rhythmic groups acquired the characteristics of figures on a background. Therefore, it was not surprising that intervals of the same duration were perceived differently when they separated two sounds that belonged to the same group and when they separated two sounds that belonged to different groups.

Similarities with the Perception of Speech

In this context, it is useful to note some similarities of the perceptual phenomena just described to some effects that have been observed in speech perception. Cue-trading relationships, that is, substituting a change in one dimension of sound for a change in another dimension that results in a perceptually indistinguishable stimulus, are common in speech perception research. The concept of accent in music is similar to the concept of stress in speech, and potential cues to stress in speech are multiple as well. Fry (1955, 1958) used words that take on different meanings with different stress patterns (e.g., "object") in some experiments. He found that either lengthening one vowel, increasing its amplitude, or increasing its fundamental frequency relative to the other vowel in the word could determine where the stress was perceived by subjects. Fry also investigated the production of sounds and found that speakers combined duration and amplitude cues in order to produce an utterance with a specified stress pattern.

In Fry's (1955, 1958) experiments, the actual perceptual equivalence of stresses created by the use of variations of different dimensions of sound was not investigated. However, such trading relationships have been observed in speech in other contexts. Fitch, Halwes, Erickson, and Liberman (1980) found the duration of the gap before the onset of voicing and the onset frequency of the first formant to be interchangeable as cues in distinguishing between the words "slit" and "split." It was possible to find pairs of gap durations and compensating formant onset frequencies such that the resulting stimuli would be very hard to distinguish even though they differed in two dimensions. Best, Morrongiello, and Robson (1981) replicated these results using the words "say" and "stay."

Another relevant observation is about the perception of the relative timing of stressed sounds in speech (Fowler, 1979; Morton, Marcus, & Frankish, 1976). When subjects perceived a sequence of spoken words as having stressed syllables separated by equal time intervals, the onsets of the stressed sounds systematically deviated from equal timing. Interstress times were longer after longer consonants than shorter consonants, just as in Woodrow's (1909) experiment with long and short tones, in which the onsets of longer tones had to be separated from the onsets of the following tones by longer intervals in order for perceptual isochrony to be observed.

Thus, considering distortions in the perception of one dimension of sound in musical stimuli as a result of a manipulation of another dimension to be assimilation into a rhythmic structure has a counterpart in speech perception: Just as tokens of the same phoneme created by different combinations of physical variables are hard to distinguish from each other within certain limits (Best et al., 1981; Fitch et al., 1980), tokens of the same rhythmic structure realized by different physical variables may also act as perceptual equivalents. The perceived rhythmic structure may lead to mistaken attributions about the actual dimensions of sound that give rise to it.

General Framework for the Experiments

The experiments reported in this article aimed to demonstrate perceptual equivalence of some dimensions of sound in establishing rhythmic patterns. If a change in one dimension can lead to the perception of a change in another dimension, it may be possible to combine changes in both dimensions such that one may reinforce or counteract the other.

How two dimensions of sound interact at the level of the perceived temporal structure was investigated in experiments analogous to the experiment by Fitch et al. (1980) with speech. Each experiment had an identification part and a discrimination part. In each trial in the identification part, subjects decided whether a sequence that they listened to had a regularly repeating pattern in it. In each trial of the discrimination part, subjects tried to discriminate two different repeating patterns from each other. The question was whether performance in the discrimination task could be predicted from the probabilities of identifying sequences as having or not having rhythmic patterns. The probability that a sequence would be perceived as having a rhythmic pattern depended on how the variations in the two dimensions were combined. Thus, two pairs of sequences that differed by the same magnitude on the same two dimensions could have varying differences in terms of the strength of their rhythmic pattern, depending on how the differences of the two dimensions were combined: The two sequences could be very different if one variable reinforced the difference that was created by the other, but the two sequences could be similar if one variable countered the effect of the other variable.

The critical distinction was that if the subjects were using the differences in the two variables individually in making discriminations, then two sequences that differed by the same amount on the same two dimensions should be equally discriminable no matter how the two differences were combined. However, if the subjects were using the presence or absence of a perceived rhythmic pattern in the two sequences, then discrimination should be better when the variations in two variables cooperate to emphasize the difference of the overall rhythmic pattern than when the same differences undermine each other's effects.

Three pairs of dimensions were investigated in three experiments, respectively: Intensity and duration of intertone intervals (Experiment 1), pitch intervals and intensity (Experiment 2), and duration of intertone intervals and pitch intervals (Experiment 3). The first member of a pair of dimensions had two levels, and the second member had seven different levels. For clarity of exposition, the terms *accenting variable* and *probe variable* will be used to refer to the former and latter dimensions, respectively. However, variations of the probe variable could result in perception of accents as well, and the subjects were not instructed to attend to the probe variable only.

Experiment 1

In Experiment 1, the accenting variable was intensity and the probe variable was the duration of the intertone intervals. Both a higher intensity tone and a longer silent interval preceding a tone can lead to the perception of the beginning of a group. If longer silent intervals preceded higher intensity tones, then the two cues were said to be "cooperating." If shorter silent intervals preceded higher intensity tones, then the two cues were said to be "conflicting." The hypothesis was that subjects would depend on the difference between the strengths of rhythmic grouping in the two sequences rather than the differences in individual dimensions. As a result, it was expected that the discrimination performance could be predicted from the identification performance. The predictions would be different for cooperating and conflicting pairs of differences, even when the magnitude of the individual differences remained the same.

METHOD

Subjects

Twenty undergraduates at Dartmouth College participated in the experiment in exchange for extra credit in an introductory psychology course. The average length of time they had played a musical instrument or performed with voice was 6 years. The range was from 0 to 15 years. Although musically trained people are generally more accurate in making perceptual judgments of the kind used in these experiments, their formal knowledge may affect their judgments. In any case, the goal of the present research was to investigate perceptions of people who have passive exposure to music and not necessarily formal instruction about it.

Stimuli

Stimuli were sequences of sinusoidal tones ascending or descending in frequency in one-semitone steps. The *unaccented* version of these sequences consisted of 21 tones, all of equal intensity and of 210-ms duration. This duration included 10-ms attack and 10-ms decay times. The tones were separated by silent intervals of 50 ms.

Accented sequences with no deviation of the probe variable were created by changing the value of regularly spaced tones in the unaccented sequences on the accenting variable (i.e., intensity). The first tone to be accented was one of the first three tones in the sequence, selected randomly for each trial. Every third tone following the first accented tone was also accented. In Experiment 1, these accents were created by an intensity that was 3 dB higher than the intensity of the unaccented tones.

Seven different versions each of the accented and unaccented sequences were created. These versions differed in terms of the deviation of the probe variable (i.e., intertone interval duration) from the standard. This deviation was used for every third tone in a sequence. In the accented sequences, the deviant intertone intervals preceded the higher intensity tones. The particular ways in which the deviant tones differed from the rest of the tones in a sequence are described in Table 1. Thus, 14 distinct types of sequences were included in the stimulus set.

Procedure

Identification

In the identification part of the experiment, the subjects were asked to decide whether or not a sequence they heard had a pattern in it. Before performing the experimental trials in the identification part, subjects heard a set of example sequences. In some of the examples, all the tones had equal intensities and all the intertone intervals were equal. In others, patterns were created by inclusion of louder tones or longer intertone intervals, but not both. The manipulations in the examples were larger than any of the variations of intensity and timing in the experimental trials. The examples were repeated if the subject failed to notice the distinction. Before starting the trials, the subjects were informed that the patterns in the experimental trials would be more subtle but they would still be variations of the examples they had heard.

The identification part included 140 trials. Each one of 14 types of sequences was replicated 10 times. Thus, the identification part had a 2 x 7 (Levels of Accenting Variable x Levels of Probe Variable) design.

The starting pitch was selected randomly for each sequence. For ascending sequences, the starting pitch was selected from among the chromatic tones between $F\sharp_3$ and B_3. For descending sequences, the starting tone was selected from among the chromatic tones between $F\sharp_4$ and B_4. Half the sequences in the identification part had ascending pitch, and the

TABLE 1
Deviations of the Probe Variable and the Corresponding Manipulations in the Sequences in Experiments 1, 2, and 3

Experiment 1		Experiment 2		Experiment 3	
Deviation of Probe Variable (ms)	Duration of Deviant Intertone Interval (ms)	Deviation of Probe Variable (dB)	Intensity Difference of Deviant Tone (dB)	Deviation of Probe Variable (cents)	Deviant Pitch Change (cents)
−20	30	−3.0	−3.0	−50	50
−10	40	−1.5	−1.5	−25	75
0	50	0.0	0.0	0.0	100
10	60	1.5	1.5	25	125
20	70	3.0	3.0	50	150
30	80	4.5	4.5	75	175
40	90	6.0	6.0	100	200

other half had descending pitch. Subjects had two response alternatives: They could either respond that there was a repeated pattern in the sequence or they could respond that there was no pattern.

Discrimination

A three-way oddity procedure was used in the discrimination part of the experiment. The oddity procedure required fewer trials than a yes/no procedure, which would have made the experimental sessions longer. In each trial, subjects heard three sequences, two of which had the same repeated pattern of intensities and intertone intervals. The other sequence, the odd one, had a different pattern. Subjects were instructed to indicate which sequence had the different pattern in each trial. Subject were expected to have a sense of the kinds of patterns from their exposure to the sequences in the identification part. None of the subjects had a problem with this instruction. Feedback was not given during the discrimination part.

The three sequences in one trial were selected randomly from among the sequences in the identification part. The direction of pitch change could be different for one sequence compared with the other two in one trial, but this difference was not correlated with the differences in the accent patterns of the sequences. Subjects were warned of this before they started doing the three-way oddity task, and they were told to make their discriminations in terms of the kinds of patterns in the sequences only.

There were three types of comparisons: One-cue, two-cues-conflicting, and two-cues-cooperating. In the one-cue comparisons, the odd sequence in a trial differed from the others in terms of the accenting variable only (i.e., the presence or absence of tones of higher intensity). In the two-cues comparisons, the odd sequences differed in terms of presence or absence of higher intensity tones and also in terms of the probe variable (i.e., the duration pattern of intertone intervals): In the two-cues-cooperating comparisons, the two differences were combined to accent the same tones more strongly in one type of sequence than in the other, whereas in the two-cues-conflicting comparisons, in each type of sequence, one cue would undermine the accents on the tones accented by the other cue.

In the two-cues-conflicting comparisons, in the sequence that contained the high-intensity tones, the intertone intervals that preceded the high-intensity tones were 20 ms shorter than those in the other type of sequence. In the two-cues-cooperating comparisons, the combination of the variations in intensity and timing was reversed so that the longer deviant intertone intervals preceded the tones with higher intensities.

For example, the one-cue comparison with 20-ms average deviation of the probe variable involved the following kinds of sequences: A sequence of equal intensity tones in which every third intertone interval was 20 ms longer than the standard (70-ms total duration) and a sequence with the same temporal structure in which the tones that followed the 70-ms intertone intervals also had higher intensity.

In the two-cues conditions, the two types of sequences that were compared had different temporal structures and different intensity structures. A two-cues-cooperating comparison with 20-ms average variation of the probe variable involved the following kinds of sequences: One sequence of equal-intensity tones in which every third intertone interval was 10 ms longer than the standard (60-ms total duration) and one sequence in which every third tone had higher intensity and also followed an intertone interval that was 30 ms longer than the standard (80-ms total duration). Thus, both the larger deviations of the intertone intervals and the presence of the higher intensity tones supported the perception of a pattern in the latter type of sequence more strongly than in the former type of sequence.

The two-cues-conflicting comparison with 20-ms average deviation of the probe variable involved the following kinds of sequences: One sequence of equal-intensity tones in which every third intertone interval was 30 ms longer than the standard (80-ms total duration) and one sequence in which every third tone had higher intensity and also followed an intertone interval that was 10 ms longer than the standard (60-ms total duration). In this type of comparison, the temporal pattern supported perception of a pattern in the former type of sequence more strongly, whereas the higher intensity tones supported the perception of a pattern in the latter type of sequence more strongly.

Five different versions of each type of comparison, differentiated by the average deviation of the intertone intervals of the two types of sequences compared, were used in the experiment. Thus, the discrimination part had a 3 x 5 (Type of Comparison x Average Deviation) factorial design. In this design, each factorial combination had six replications for each subject. Which type of sequence in a comparison was the odd one and the serial position in which the odd sequence was heard in the trial were counterbalanced across the six replications. Thus, the discrimination part consisted of 90 trials. All types of trials were mixed randomly in one session.

RESULTS AND DISCUSSION

Identification

Average percentages of responses that indicated that the subject thought there was a pattern in each type of sequence are presented in Figure 1. Analysis of these data revealed significant main effects of relative intensities [$F(1,19) = 60.02$, $MS_e = 815$, $p < .001$] and relative timing [$F(6,114) = 42.26$, $MS_e = 407$, $p < .001$] on the probability that a type of sequence would be classified as having a repeated pattern in it. The interaction of these two factors did not reach significance at the .05 level [$F(6,114) = 1.95$, $MS_e = 280$, $p = .079$].

The probabilities of classifying each type of sequence as having a repeated pattern in it was used to calculate predicted probabilities of responding correctly in the three-way oddity test. The assumption was that two sequences would be perceived as being different only if one was perceived

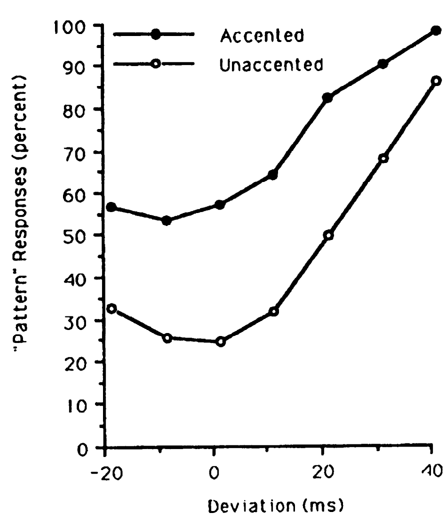

Fig. 1. Average percentages of trials in which subjects reported hearing a repeated pattern for each type of sequence in the identification part of Experiment 1.

as having a pattern and the other was perceived as not having a pattern. If subjects used the identification of a repeated pattern as the basis for discrimination in the three-way oddity task, accuracy in the three-way oddity task should be accurately predicted by the following formula, which gives the probability that the odd sequence will be classified differently than the two sequences with the same pattern (MacMillan & Creelman, 1991, p. 173):

$$P_c(s_1,s_2) = \{1 + 2 \times [P(s_1) - P(s_2)]^2\}/3,$$

where $P_c(s_1,s_2)$ is the probability of responding correctly in a discrimination trial that involves a comparison between sequences s_1 and s_2, and $P(s_1)$ and $P(s_2)$ are the probabilities that sequences s_1 and s_2, respectively, were identified as having a repeated pattern by that subject.

This formula led to the prediction that performance in the two-cues-cooperating, one-cue, and two-cues-conflicting conditions should proceed from best to worst in that order. However, if the differences in the two variables were used independently in making the discrimination decision, there was no reason to expect that the performance in the two-cues-cooperating and the two-cues-conflicting conditions should have been different. In that case, the probability of making a correct discrimination would be equal to the probability of detecting at least one of the differences correctly.

Discrimination

The average percentages of predicted and observed correct discrimination responses are presented in Figure 2. The task was expected and observed to be rather difficult for the subjects. The performance in some of the conditions was not reliably better than chance. These conditions were one-cue comparisons with average deviations of the probe variable of –10, 20, and 30 ms, two-cues-conflicting comparisons with average deviations of the probe variable of 0 and 30 ms, and two-cues-cooperating comparisons with an average deviation of the probe variable of 0 ms.

A two-way analysis of variance (ANOVA) conducted on the observed discrimination data revealed a main effect of comparison type [$F(2,38) = 6.57$, $MS_e = 276.66$, $p < .005$]. The average percentages of correct responses for the one-cue, two-cues-conflicting, and two-cues-cooperating comparisons were 44.33%, 43.33%, and 51.17%, respectively. Tukey's HSD for these three means was 5.82%. Thus, the performance in the two-cues-cooperating condition was better than the performance in the one-cue and the two-cues-conflicting conditions, which were not significantly different from each other. The effect of adding the second cue was slightly less than predicted. The corresponding predicted average percent-

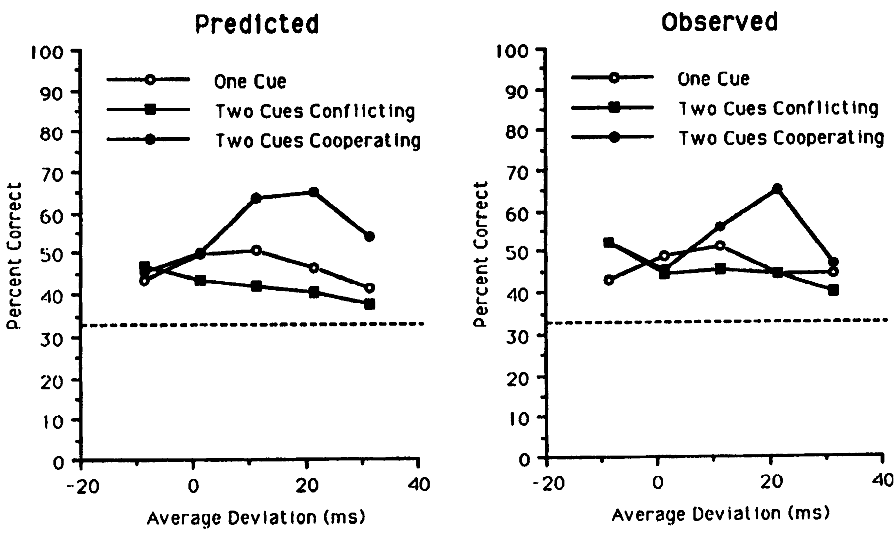

Fig. 2. Predicted and observed percentages of correct responses in the discrimination part of Experiment 1. Average deviation refers to the mean deviations of the durations of the deviant intertone intervals in the two sequences compared in one trial. Deviation is the difference of the duration of the intertone intervals from the standard duration of 50 ms.

ages of correct responses were 44.43%, 40.28%, and 53.87%, respectively. The observed main effect of the average deviation of intertone intervals [$F(4,76) = 1.71$, $MS_e = 354.46$, $p = .155$] and the interaction of the two factors [$F(8,152) = 1.62$, $MS_e = 392.4$, $p = .122$] were not significant, although the predicted discrimination performance suggested significant effects.

A similar ANOVA was conducted on the difference scores (observed minus predicted percentages). This analysis indicated that the observed data did not systematically differ from the predictions based on the assumption that whether or not a sequence had a noticeable rhythmic pattern was the information used in making the discriminations. The average difference was not significantly different from zero [$F(1,19) < 0.01$]. The main effects of type of comparison [$F(2,38) = 2.41$, $p = .103$], average deviation of intertone intervals [$F(4,76) = 0.79$, $p = .536$], and the interaction of the two factors [$F(8,152) = 0.63$, $p = .755$] were not significant.

To summarize, it was possible to predict discrimination performance with reasonable accuracy from the identification performance. Most importantly, performance in the two conditions that involved two cues differed from each other in a way that was predicted from the identification data. This showed that the differences in the two variables did not make independent contributions to discrimination performance.

As another example of the accuracy of the predictions from the identification data, performance in the two-cues-cooperating condition declined with the largest average deviation of the probe variable. Because the accented sequences with larger deviations of the probe variable were almost always perceived as having rhythmic patterns, the difference between the probabilities for pairs of sequences in two-cues-cooperating condition decreased. This led to prediction of a decline in accuracy for this type of comparison with larger deviations of the probe variable, which was confirmed by the observed data.

Experiment 2

In Experiment 2, the accenting variable was pitch and the probe variable was intensity. Both a larger change in the pitch and a higher intensity tone can lead to the perception of a beginning of a group. If the larger pitch change ended in a higher intensity tone, then the two cues were said to be "cooperating." If a larger change in pitch ended in a lower intensity tone, then the two cues were said to be "conflicting." The design and the reasoning behind it were analogous to those of Experiment 1.

METHOD

Subjects

Twenty Dartmouth College students participated in the experiment in return for extra credit in an introductory psychology course. The average length of time that they had played musical instruments or performed with voice was 5.4 years. The range was from 0 to 13 years.

Stimuli

Sequences of 18 tones[1] similar to those used in Experiment 1 were used as stimuli. Accented sequences were created by pitch skips of three semitones,[2] rather than the standard chromatic scale step of one semitone, before every third tone. The intensity of every third tone was varied by steps of 1.5 dB across sequences. Thus, seven variations each of accented and unaccented sequences in which the intensity of every third tone varied from 3 dB less than the standard to 6 dB more than the standard were created (Table 1).

1. In Experiments 2 and 3, the sequences were reduced from 21 tones to 18 tones. This was done to reduce the duration of the experimental session, which appeared to be uncomfortably long for the subjects in Experiment 1.

2. Using a pitch change of two semitones, rather than three semitones, was tried but pilot subjects did not show any indication of discriminating between sequences that included steps of two semitones and those that included semitone steps only. For this reason, pitch changes of three semitones were used as the means of accenting, although this turned out to be noticed too easily by the subjects.

Procedure

The procedure was analogous to the procedure of Experiment 1. In the discrimination part, when the comparison involved a difference in relative intensities, the deviation in one type of sequence was 3 dB larger than the deviation in the other type of sequence. The two-cues-conflicting comparisons were created by combining larger changes in pitch with tones of lower intensities and the two-cues-cooperating comparisons were created by combining larger changes in pitch with tones of higher intensities. Because of the larger pitch range covered by the sequences in this experiment, the beginning tones of the ascending sequences were selected from the F_3-$B\flat_3$ range, and the starting tones of the descending sequences were selected from the B_5-E_6 range.

RESULTS AND DISCUSSION

Identification

The average percentages of responses that indicated that subjects detected repeated patterns in the different types of sequences are presented in Figure 3. An ANOVA indicated a significant effect of pitch structure [$F(1,19) = 29.71$, $MS_e = 2030$, $p < .001$] and relative intensities [$F(1,19) = 74.92$, $MS_e = 227$, $p < .001$]. The probability of detecting a pattern increased with the magnitude of the difference in the intensities of the tones in a sequence. The introduction of the pitch skips of three semitones greatly increased the probabilities of detecting patterns. This effect was modified by the interaction of the two factors [$F(1,19) = 16.79$, $MS_e = 219$, $p < .001$]. This seemed to be due to a ceiling effect, which resulted in the decrease of the effect of

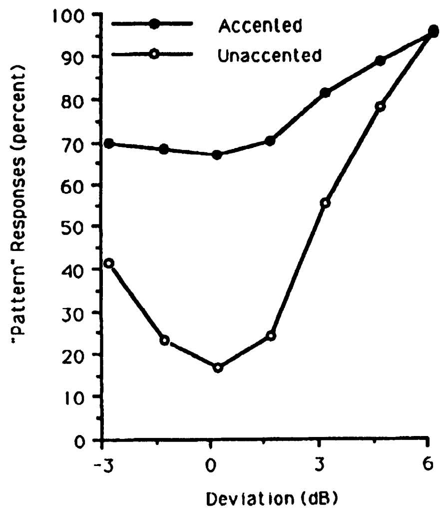

Fig. 3. Average percentages of trials in which subjects reported hearing a repeated pattern for each type of sequence in the identification part of Experiment 2.

the pitch structure for sequences that were almost always classified as having a pattern by virtue of the large intensity differences involved.

Discrimination

The predicted average percentages of correct responses in the discrimination part were calculated in the same way as in Experiment 1. The predicted and the observed averages are presented in Figure 4. The discrimination performance in one-cue comparisons with –1.5-dB average deviation of the probe variable and two-cues-conflicting comparisons with 4.5-dB average deviation of the probe variable were not reliably different from chance.

For the observed data, the main effect of the type of comparison was close to significance [$F(2,38) = 3.04$, $MS_e = 452.78$, $p = .059$]. The observed average percentages of correct responses for one-cue, two-cues-conflicting, and two-cues-cooperating comparisons were 53.46%, 53.94%, and 60.14%, respectively. The corresponding predicted averages were 50.14%, 47.42%, and 56.46%. The interaction of the type of comparison and the average deviation of intensities in the sequences was also significant [$F(8,152) = 3.14$, $MS_e = 337.89$, $p < .005$].

Although the observed performance was better than the predicted performance, the analysis of the difference scores (observed minus predicted

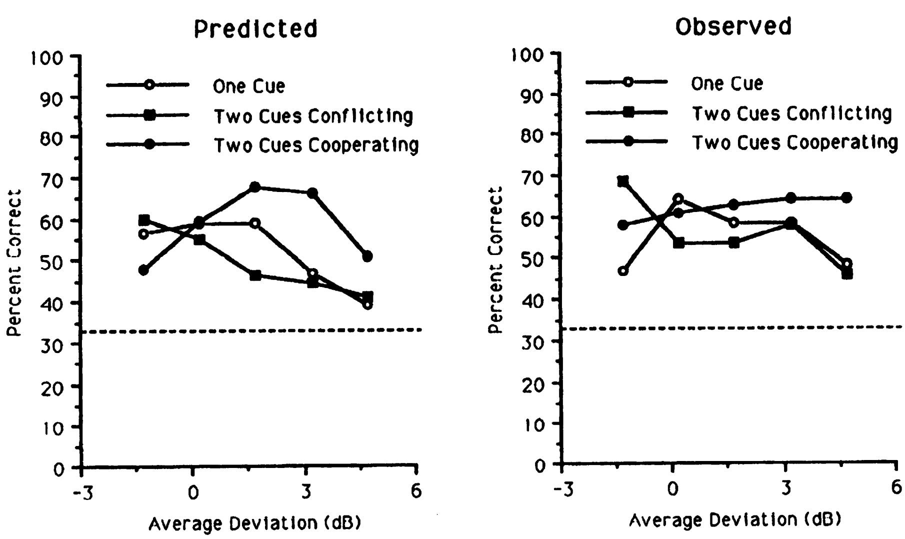

Fig. 4. Predicted and observed percentages of correct responses in the discrimination part of Experiment 2. Average deviation refers to the mean deviations of the intensities of the deviant tones in the two sequences compared in one trial. Deviation is the difference in the intensity of the tones from the standard intensity.

percentages) did not show this difference to be statistically significant [$F(1,19) = 3.05$, $MS_e = 2005.88$, $p = .097$]. For the difference scores, the main effects of type of comparison [$F(2,38) = 0.61$, $p = .548$] and the average deviation of intensities [$F(4,76) = 1.82$, $p = .134$] were not significant either. However, the interaction of the two factors was significant [$F(8,152) = 3.14$, $MS_e = 381.17$, $p < .005$]. This indicated that the deviations from the predicted scores differed among the 15 means involved. Underestimation of discrimination performance in some conditions such as two-cues-cooperating with 4.5-dB average deviation [13.8%, $t(152) = 3.16$, $p = .002$], two-cues-conflicting [13.1%, $t(152) = 3.0$, $p = .0032$] and one-cue with 3.0-dB average deviation [11.5%, $t(152) = 2.63$, $p = .0091$] could be due to the subjects' ability to make use of the fact that the pitch structures of the two sequences were different even though they were similar on the basis of having strong accent structures if the type of accents was not considered. This was not surprising, as the change in pitch structure was observed to be very noticeable to the subjects in the identification part.

The observed pattern of performance in the discrimination part had a close similarity to the predicted performance. Nevertheless, there were signs that in some cases subjects did use the specific differences in making discriminations rather than making use of presence or absence of accents in the sequence regardless of the means by which they were created. This may well be because the magnitude of the pitch difference was too large to be effectively compensated for by the change in the intensity. The accuracy of the predictions based on the identification performance may have decreased because of a possible ceiling effect on the identification performance.

Another remarkable result was the reversal of the difference between the two-cues-cooperating and the two-cues-conflicting conditions with the –1.5-dB average deviation relative to the larger average deviations. This reversal was predicted from the identification performance and observed in the discrimination performance. This switch-over indicates that the cue trading was not at the level of tones but at the level of sequences. In other words, it was not the case that louder tones that came after small changes in pitch were confused with softer tones that came after larger changes in pitch. This would result in better performance in the two-cues-cooperating condition for all average deviations of the intensities. Rather, sequences that had rhythmic patterns with similar strengths were confused with each other even though the accents were created by different means.

Experiment 3

Experiment 3 was analogous to Experiments 1 and 2. The accenting variable was the duration of the intertone intervals and the probe variable was the pitch intervals. If longer intertone intervals coincided with larger changes in pitch, then the two cues were said to be cooperating. If longer intertone intervals coincided with smaller changes in pitch, then the two cues were said to be conflicting.

METHOD

Subjects

Twenty Dartmouth College students participated in the experiment in return for extra credit in an introductory psychology course. The average length of time that they had played a musical instrument or performed with voice was 6.4 years. The range was from 0 to 14 years.

Stimuli

Sequences similar to those used in Experiments 1 and 2 were used. The accents were created by intertone intervals of 80 ms, rather than the standard 50 ms, before every third tone in the accented sequences. Pitch was varied by steps of 25 cents (1/4 semitones): Every third tone in a sequence was shifted in terms of pitch toward (positive deviations) or away from (negative deviations) the tone that followed it by a multiple of 25 cents. The manipulation of every third pitch interval was compensated for by a complementary change in the following pitch interval, such that the two intervals added up to two semitones. Thus, when a tone was moved closer in pitch to the preceding tone it was moved farther from the following tone, and vice versa. This strategy was chosen because manipulating the pitch by such small intervals without compensation proved to be too difficult for the subjects to detect. A description of the seven variations each of accented and unaccented sequences that resulted can be found in Table 1.

Procedure

The procedure was identical to the procedure of Experiments 1 and 2. In the comparisons that involved differences in pitch structure, the two types of sequences differed in terms of deviation by 50 cents. The two-cues-conflicting comparisons were created by juxtaposing longer intertone intervals with smaller changes in pitch, and the two-cues-cooperating comparisons were created by juxtaposing longer intertone intervals with larger changes in pitch.

RESULTS AND DISCUSSION

Identification

The average percentages of trials in which the subjects reported detecting a repeated pattern in the sequences are presented in Figure 5. A two-way ANOVA indicated that manipulations of both intertone intervals [$F(1,19) = 121.66$, $MS_e = 191$, $p < .001$] and pitch intervals [$F(6,114) = 37.79$, $MS_e = 379$, $p < .001$] significantly affected the probability that a repeated pattern would be detected. The interaction of these two factors was also significant [$F(6,114) = 5.95$, $MS_e = 191$, $p < .001$]. Unlike the case in Experiments 1 and 2, accenting increased the probability of a pattern being detected more if it was coupled with a negative deviation rather than a positive deviation of the probe variable.

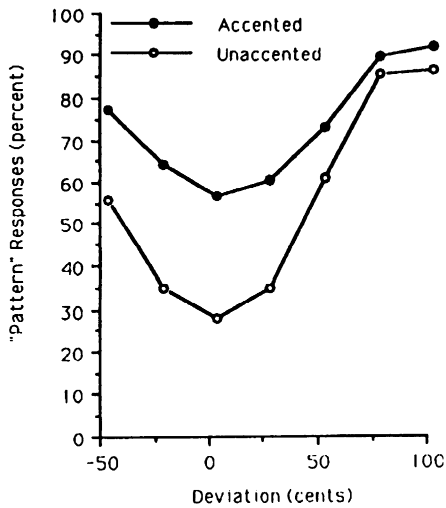

Fig. 5. Average percentages of trials in which subjects reported hearing a repeated pattern for each type of sequence in the identification part of Experiment 3.

Discrimination

The predicted average percentages of correct responses in the discrimination part were calculated in the same way as in Experiments 1 and 2. The predicted and observed percentages are presented in Figure 6. The performance in one-cue and two-cues-conflicting comparisons with 75-cents average deviation of the probe variable was not reliably better than chance.

The analysis of the observed performance found the main effects of comparison type [$F(2,38) = 1.53$, $MS_e = 532.31$, $p = .231$] to be nonsignificant. The average percentages of correct responses for one-cue, two-cues-conflicting, and two-cues-cooperating comparisons were 47.5%, 52.17%, and 52.67%, respectively. These averages were higher than their predicted counterparts, which were 39.21%, 42.85%, and 46.58%, respectively. The main effect of the average deviation of pitch intervals was on the verge of significance [$F(4,76) = 2.49$, $MS_e = 567.45$, $p = .05$] for the observed data. This was due to lower discrimination performance in the trials with an average deviation value of 75 cents. The analogous effect was found in the predicted data. The interaction of the two factors was significant [$F(8,152) = 2.29$, $MS_e = 326.11$, $p < .05$].

The analysis of the difference scores (observed minus predicted percentages) showed that the observed performance was significantly better than the predicted performance [$F(1,19) = 19.22$, $MS_e = 973.98$, $p < .001$]. The main effects of the type of comparisons [$F(2,38) = 0.43$, , $p = .654$], the

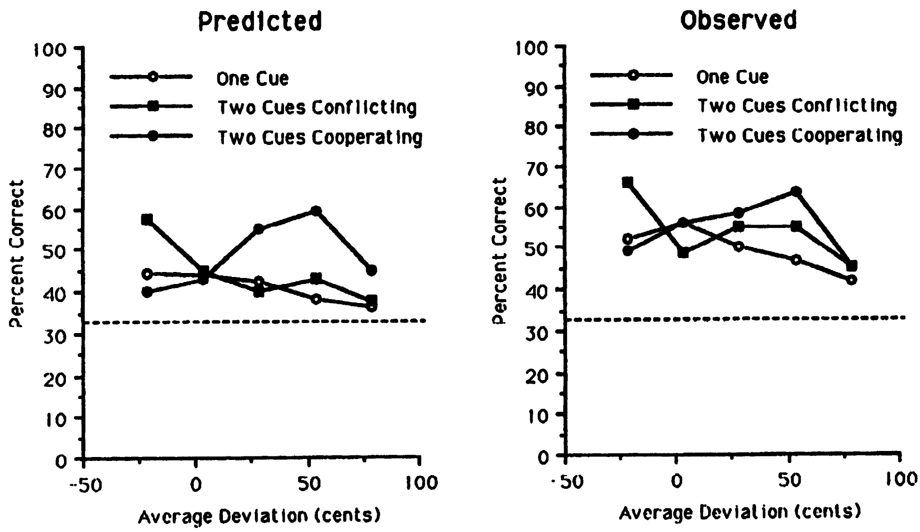

Fig. 6. Predicted and observed percentages of correct responses in the discrimination part of Experiment 3. Average deviation refers to the mean deviations of the deviant pitch intervals in the two sequences compared in one trial. Deviation is the difference of the interval from the standard interval of one semitone.

average deviation of pitch intervals [$F(4,78) = 0.44$, $p = .777$], and the interaction of the two factors [$F(8,152) = 0.95$, $p < .479$] were not significant.

The results of Experiment 3 were equivocal. A significant effect of the type of comparison was not found. The average percentage of correct responses for two-cues-conflicting was almost equal to that for two-cues-cooperating comparisons, and both were slightly higher than the average percentage of correct responses for the one-cue comparison. However, analysis of the difference scores did not indicate a statistically significant deviation from the pattern of the predicted discrimination performance. In general, however, observed discrimination performance was better than predicted. This may mean that in the discrimination part, subjects attended to differences to which they did not pay as much attention in the identification part.

As in Experiment 2, the nature of the comparison type by average deviation interaction in the observed data was similar to the nature of the predicted interaction. Two-cues-conflicting comparisons were easier than two-cues-cooperating comparisons for the –25 cents average deviation, as predicted. This result converges with the analogous result from Experiment 2 that subjects were more likely to confuse sequences that had similar accent structures than to confuse sequences that differed greatly in terms of the presence or absence of accents, regardless of what type of accents were involved.

General Discussion

In the three experiments reported, it was found that performance in a discrimination task involving sequences of tones with different accent structures could be predicted to a large extent from the probabilities that each type of sequence would be perceived as having a regular pattern in it. The specific type of manipulation was found to have an effect on discrimination performance in some conditions of Experiment 2. This could be due to the large magnitude of the pitch manipulation used in that experiment. The three-semitone skip that was used might have been too large to be compensated for by a change in another variable.

However, one critical prediction about discrimination performance based on identification performance, namely, that discrimination in the two-cues-cooperating conditions should be better than discrimination in the two-cues-conflicting conditions, was not borne out in Experiment 3. This may be because pitches are perceived categorically along a chromatic scale (Burns & Ward, 1978), so that nonchromatic variations of pitch may not interact with other dimensions of sound in the way that continuous dimensions such as time and intensity do.

An attempt to explain these trading relationships on the basis of lower level interactions (e.g., higher intensity tones making the silent intervals that precede them appear longer than they actually are because of some characteristic of the sensory system) would fail in the face of the interaction of comparison type with average deviation. Such a sensory effect would be expected to be the same regardless of the average deviation that is involved in the comparison. Especially problematic for this kind of explanation is the reversal of the performances in the two-cues-conflicting and the two-cues-cooperating conditions in Experiments 2 and 3, which were predicted from the identification data. If what made the two types of sequences in the two-cues-conflicting comparisons sound similar was a lower level interaction of the two variables manipulated in the experiment, the two-cues-cooperating comparisons should have been easier than the two-cues-conflicting comparisons for all levels of average deviations. This finding can be more consistently explained if discrimination is assumed to rely on the identification of a rhythmic pattern in the sequences.

The results support the view that in perception of temporal patterns, the stimuli are assimilated into some mental framework. This view has been proposed as an explanation of how people perceive and reproduce temporal patterns made up of varying durations (Fraisse, 1956, 1982, 1984; Jones, 1976; Jones & Boltz, 1989; Povel, 1981). It was found that such patterns were distorted in a way that could be described as fitting onto a sequence of isochronous beats, like an internal clock. Similarly, in the present experiments, the qualities of accented sounds are perceived as a package that involves variations in several dimensions, even if the physical stimuli may not support all of these variations.

Furthermore, the perception of the accent rather than the perception of the individual dimensions that give rise to it appears to be primary. Within certain limits, people may be aware of the presence of rhythmic patterns that result from separation by time, pitch, or intensity when they are not aware of which specific factor is responsible for the effect. This can contribute to the changes in the perception of one dimension of sound that result from the manipulation of another dimension (Povel & Okkerman, 1981; Tekman, 1993; Woodrow, 1909). The present results are congruent with other research I have done (Tekman, 1993), in the sense that cue trading was not unambiguously observed in Experiment 3, in which timing and pitch were the accenting and the probe variables, respectively. Changing timing was not found to change the perception of pitch intervals either.

A natural extension of this research, using other classifications of tone sequences, could yield clearer answers to the questions investigated in this research. An alternative classification that may be used in order to investigate cue trading in perception of rhythmic patterns may be contrasting different rhythms, such as duple and triple, rather than presence or absence of rhythmic patterns. Using stimuli on a continuum from unambiguously duple rhythm to unambiguously triple rhythm, one can attempt to predict discrimination performance from identification performance. Use of this procedure with manipulations of a single variable can clarify whether rhythmic structure is perceived categorically. Use of the same procedure with manipulations of two variables may bring more conclusive answers to the question of whether cue trading is possible in the perception of rhythmic structure.[3]

References

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics, 29,* 191–211.

Burns, E. M., & Ward, W. D. (1978). Categorical perception—phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America, 63,* 456–468.

3. The experiments reported in this article constitute part of a doctoral dissertation submitted to the faculty of Dartmouth College in June 1992. The examining committee for the dissertation consisted of Carol Fowler, John Jalowiec, Larry Polansky, George Wolford, and Jamshed Bharucha. The research was supported in part by Grant BNS-8910778 to Jamshed Bharucha from the National Science Foundation. I thank two anonymous reviewers for their comments on an earlier version of this article.

Fitch, H. L., Halwes, T. G., Erickson, D. M., & Liberman, A. M. (1980). Perceptual equivalence of two acoustic cues for stop consonant manner. *Perception & Psychophysics, 27,* 343–350.

Fowler, C. A. (1979). "Perceptual centers" in speech production and perception. *Perception & Psychophysics, 25,* 375–388.

Fraisse, P. (1956). *Les structures rythmiques [Rhythmic structures].* Louvain: Publications Universitaires de Louvain.

Fraisse, P. (1978). Time and rhythm perception. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 8). New York: Academic Press.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.) *The psychology of music.* New York: Academic Press.

Fraisse, P. (1984). Perception and estimation of time. *Annual Review of Psychology, 35,* 1–36.

Friberg, A. (1991). Generative rules for music performance: A formal description of a rule system. *Computer Music Journal, 15,* 56–71.

Friberg, A., Frydén, L., Bodin, L. G., & Sundberg, J. (1991). Performance rules for computer-controlled contemporary keyboard music. *Computer Music Journal, 15,* 49–55.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustic Society of America, 27,* 765–768.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech, 1,* 126–152.

Gabrielsson, A. (1974). Performance of rhythm patterns. *Scandinavian Journal of Psychology, 15,* 63–72.

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. (1992). *The structure and function of accent patterns.* Paper presented at the Second International Conference for Music Perception and Cognition, Los Angeles, California.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96,* 459–491.

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

MacMillan, N. A., & Creelman, C. D. (1991). *Detection theory: A users guide.* Cambridge: Cambridge University Press.

Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review, 83,* 405–408.

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

Povel, D. J., & Okkerman, H. (1981). Accents in equitone sequences. *Perception & Psychophysics, 30,* 565–572.

Sloboda, J. A. (1983). The communication of musical meter in piano performance. *Quarterly Journal of Experimental Psychology, 5A,* 377–396.

Sundberg, J., Friberg, A., & Frydén, L. (1991). Threshold and preference quantities of rules for music performance. *Music Perception, 9,* 71–92.

Tekman, H. G. (1993). Perception of rhythmic structure in tone sequences: Interactions of relative timing, intensity, and pitch (Doctoral dissertation, Dartmouth College, 1992). *Dissertation Abstracts International, 53,* 3817B.

Tenney, J. (1964/1988). *META + Hodos and META Meta + Hodos.* Oakland: Frog Peak Music.

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

Thomassen, J. M. (1982). Melodic accents: Experiments and a tentative model. *Journal of the Acoustical Society of America, 71,* 1596–1604.

Woodrow, H. (1909). A quantitative study of rhythm. *Archives of Psychology, 14,* 1–66.

Woodrow, H. (1951). Time perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley.