Proceedings of the 11th International Conference on Music Perception and Cognition (ICMPC11). Seattle, Washington, USA.
S.M. Demorest, S.J. Morrison, P.S. Campbell (Eds)

THE PERCEPTUAL GROUPING OF MUSICAL SEQUENCES: PITCH AND TIMING AS COMPETING CUES

Kamil Hamaoui and Diana Deutsch

University of California, San Diego

ABSTRACT

This study investigated the effects of pitch and timing on perceptual grouping, and examined the relative strengths of grouping based on pitch and temporal proximity. Sequences of twelve tones were constructed in which pitch proximity suggested one type of grouping (e.g., four groups of three tones each) and temporal proximity suggested an opposing type of grouping (in this case, three groups of four tones each). Sequences were presented that varied in the magnitudes of the pitch and temporal cues, and listeners indicated for each sequence whether they heard groupings of three or of four tones. It was found that the larger the pitch distance between groups, the stronger the perceptual grouping based on pitch. However, even in sequences with groups separated by large pitch distances, grouping occurred in accordance with the temporal cue at remarkably small values. It was also found that hierarchical pitch structure had a powerful effect on perceptual grouping, even in sequences where pitch proximity was also present as a cue. The implications of these findings are discussed.

1. INTRODUCTION

A number of principles govern the perceptual grouping of patterns. Among these, the principle of proximity has proven to be very strong, and in the auditory domain can be applied to both timing (Deutsch, 1980; Dowling, 1973; Povel & Okkerman, 1981) and pitch patterns (Bregman, 1990; Deutsch, 1999; Van Noorden, 1975). Research on this issue has focused heavily on stream segregation - in which a single sequence of tones is perceived as divided into two or more separate and parallel sequences. However, in a sequence that is perceived as a single auditory stream, such as a musical line, certain tones are heard as belonging more to each other than to surrounding tones, and cluster together so as to form perceptual groupings. What factors govern this within-stream grouping process? When multiple grouping cues are present, how do the various cues compare with one another in strength of influence on perceptual grouping?

In studies that have investigated grouping in natural musical contexts, temporal proximity has often emerged as a powerful cue (Deliège, 1987; Deutsch, 1980; Frankland & Cohen, 2004; Handel, 1973). For example, Deutsch (1980) found that musically trained listeners had better memory for sequences that were temporally segmented in accordance with their pitch structure than sequences that were temporally segmented in conflict with their pitch structure. These studies, however, did not assess the relative strength of cues since the magnitudes of the cues were not systematically manipulated and varied in small increments.

The present study investigated the influence of pitch factors on perceptual grouping by varying the magnitude of an opposing temporal cue in small increments. In Part I, an experiment was conducted in which grouping based on pitch proximity was pitted against grouping based on temporal proximity. In Part II, the relationship between pitch and temporal proximity was explored in four additional experiments. In Part III, the influence of hierarchical pitch structure as a grouping cue was examined.

2. PART 1: EXPERIMENT 1

2.1 Method

Participants. Four undergraduates with normal hearing participated and were paid for their services. Two males (coded CW and RA), age 19 and 23, had 8 and 4 years of private lessons on a musical instrument, respectively; two females (coded JW and DR), both age 19, had 8 and 18 years of private lessons.

Stimuli. Sequences consisted of 12 tones each. In the default condition (i.e., no temporal cue present) the tones were 200 ms in duration and were separated by 100-ms silent intervals. Twelve pitch sequences were created using the 12-tone chromatic scale as the starting material. To create sequences consisting of groups of three tones based on pitch proximity, larger pitch distances were inserted after every three tones along the sequence. To create sequences with groups of four tones based on pitch proximity, larger pitch distances were inserted after every four tones along the sequence. Pitch distance values of 2 semitones, 5 semitones and 11 semitones were used. For each value of pitch distance, there were two types of pattern: ascending and descending. The descending pattern consisted of the same pitches as the ascending pattern, but in reverse order.

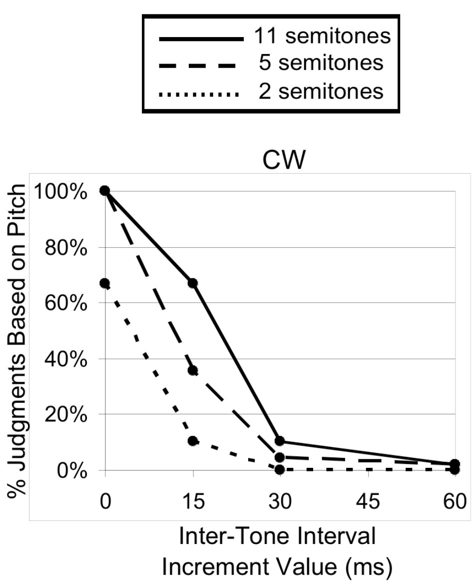
Each of the twelve pitch sequences was presented in four temporal configurations, three of which suggested a grouping contrary to that suggested by pitch proximity. In sequences with groups of four tones based on pitch proximity, the silent interval following every third tone was increased, while the silent intervals following the preceding two tones were decreased so as to keep the duration of the full pitch sequence constant. In sequences with groups of three tones based on pitch proximity, the silent interval following every fourth tone was increased with compensatory decreases in the silent intervals following the preceding three tones. The inter-tone intervals following every third or fourth tone along the sequence were increased by values of 0, 15, 30, and 60 ms. The combination of twelve types of pitch sequence with four values of the temporal cue resulted in 48 different sequences.

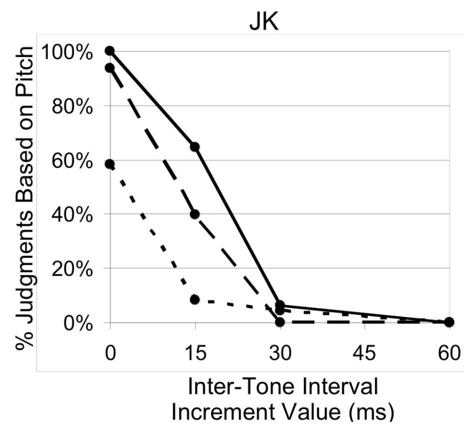
Figure 1: Examples of two sequences used in Experiment 1. The top sequence is an ascending sequence with pitch groups of three tones separated by pitch distances of two semitones. The bottom sequence is an ascending sequence with pitch groups of four tones separated by pitch distances of eleven semitones.

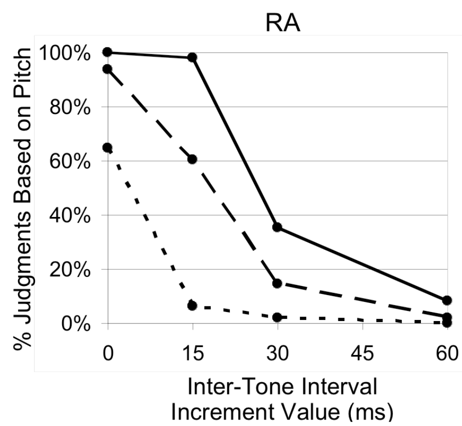
Procedure. Participants were tested individually. Each of the 48 sequences was presented twelve times over the course of four experimental sessions. Participants were instructed to judge whether they heard the tones in each sequence as forming groupings of three or of four tones each. The different types of sequence were ordered haphazardly within blocks of 12 sequences, with the constraints that no more than two sequences with the same value of temporal cue, pitch distance, or pitch group occurred in succession.

2.2 Results and Discussion

For each of the four participants, data were summed across pitch groups of three and four and also across ascending and descending sequences, resulting in 48 grouping judgments in each pitch distance/ITI increment condition. Figure 2 displays, for each participant, the percentage of groupings based on pitch in each pitch distance condition, at each value of temporal cue.







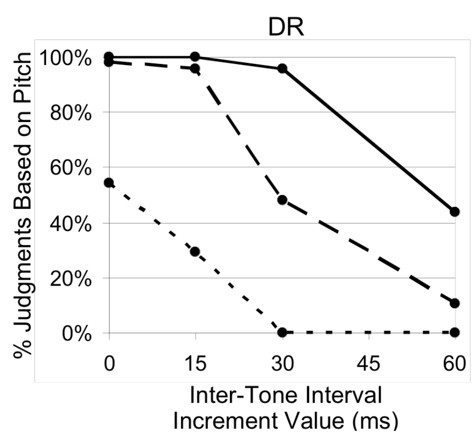


Figure 2: Individual subject data from Experiment 1.

To examine whether sequences with larger pitch distances between groups led to significantly more groupings based on pitch than did sequences with smaller pitch distances between groups, pairs of pitch distance conditions were compared at ITI increment values of 15 ms and 30 ms: Sequences with 11-semitone pitch distances between groups (11S sequences) were compared to sequences with 5-semitone pitch distances between groups (5S sequences), and 5S sequences were compared to sequences with 2-semitone pitch distances between groups (2S sequences). Comparisons were conducted using a bootstrap procedure programmed and executed with Resampling Stats software (Simon & Bruce, 1993).

For CW and JK, there were statistically significant differences between pitch distance conditions at 15 ms ($p < .01$ for all comparisons). For RA, there were statistically significant differences between pitch distance conditions at both 15 ms and 30 ms ($p < .05$ for all comparisons). For DR, there were statistically significant differences between pitch distance conditions at 30 ms ($p < .001$ for both comparisons) and between 5S and 2S sequences at 15 ms ($p < .001$). These findings indicate that the larger the pitch distance between groups within a sequence, the stronger the perceptual grouping—that is, the larger the conflicting temporal cue needs to be to compete with grouping based upon pitch proximity.

3. PART 2: EXPERIMENTS 2-5

3.1 Method

Participants. The same four participants were tested in these experiments as in Experiment 1.

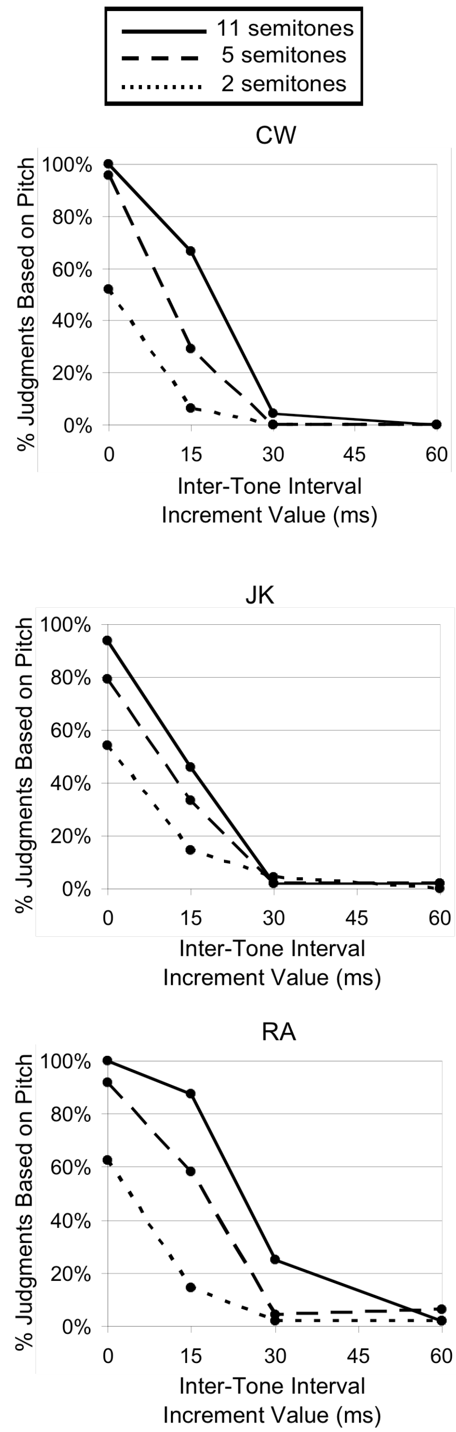
Stimuli. All sequences in Experiments 2-5 consisted of 12 tones each, using the 12-tone chromatic scale as the starting material. In Experiments 2-5, a different temporal cue was used. Increases in the silent intervals following the last tone in each group were not compensated for by decreases in the silent intervals following the previous tones in the group. In Experiment 2, the sequences were identical to those used in Experiment 1 except for the type of temporal cue used. In Experiment 3, an expanded range of temporal cue values was used: 0, 30, 60, and 120 ms. In Experiment 4, tone durations were reduced from 200 ms to 100 ms and the inter-tone intervals, in the default condition, were reduced from 100 ms to 50 ms, in effect leading to a doubling of the tempo. In Experiment 5, pitch distance values of 3, 7, and 11 semitones, and temporal cue values of 10, 20, 30, and 40 ms, were selected.

Procedure. In Experiments 2-4, the procedure was identical to that in Experiment 1. In Experiment 5, each of the 60 different sequences was presented twelve times over the course of four experimental sessions, in 12 blocks of 15 sequences each.

3.2 Results and Discussion

As in Experiment 1, data were summed across pitch groups of three and four, and across ascending and descending sequences, and comparisons of the percentage of groupings based on pitch were conducted between pitch distance conditions at the intermediate values of the temporal cue using the bootstrap procedure.

In Experiment 2, the pattern of differences for each participant was nearly identical to that in Experiment 1. Figure 3 displays, for each participant, the percentage of groupings based on pitch in each pitch distance condition, at each value of temporal cue.



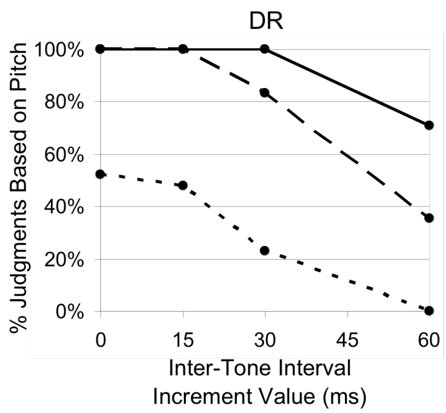


Figure 3: Individual subject data from Experiment 2.

For CW, there were statistically significant differences in the percentage of groupings based on pitch at 15 ms ($p < .005$ for both comparisons). For JK, there were statistically significant differences between 5S and 2S sequences at 15 ms ($p < .05$). For RA, there were statistically significant differences at 15 ms ($p < .005$ for both comparisons) and between 11S and 5S sequences at 30 ms ($p < .005$). For DR, there were statistically significant differences between 5S and 2S sequences at 15 ms ($p < .001$) and between pitch distance conditions at 30 ms ($p < .005$ for both comparisons). As in Experiment 1, participants CW and JK were most responsive to the temporal cue, with this cue dominating their grouping judgments for all pitch distance conditions at a value of 30 ms.

In Experiment 3, participants did not adjust their criterion for making a judgment based on the temporal cue in accordance with the expanded range of temporal cues that were used. For each participant, the temporal cue dominated grouping judgments at the same value as in Experiments 1 and 2. The temporal cue dominated grouping judgments for CW and JK at an ITI increment value of 30 ms, and for RA at a value of 60 ms ($p > .05$ for all comparisons between pitch distance conditions). For DR, however, there were differences in the percentage of groupings based on pitch between pitch distance conditions at 30 ms and 60 ms ($p < .001$ for three of the four comparisons), just as there were in Experiments 1 and 2. These findings demonstrate that the trade-off between pitch and temporal proximity is not strongly impacted by the range of temporal cue values employed.

In Experiment 4, the finding that larger pitch distances between groups lead to stronger perceptual groupings was replicated. For CW, there were statistically significant differences between 11S and 5S sequences at 15 ms ($p < .001$). For JK, there were statistically significant differences between 11S and 5S sequences at 15 ms ($p < .01$). For RA, there were statistically significant differences at 15 ms ($p < .001$ for both comparisons) and between 11S and 5S sequences at 30 ms ($p < .05$). For DR, there were statistically significant differences between 5S and 2S sequences at 15 ms ($p < .001$) and between pitch distance conditions at 30 ms ($p < .001$ for both comparisons). Similar effects of specific values of the temporal cue were revealed at the faster tempo, suggesting that the strength of the temporal cue operates on an absolute scale rather than a scale relative to the tone durations and inter-onset intervals.

In Experiment 5, the general finding was once again replicated and the data revealed that temporal cue values spaced in 10-ms steps can have differential effects on listeners' grouping judgments. For CW and RA, there were statistically significant differences between 7S and 3S sequences at 10 ms ($p < .001$ for both participants) and between pitch distance conditions at 20 ms ($p < .05$ for both comparisons for both participants). For JK, there were statistically significant differences between 7S and 3S sequences at 10 ms ($p < .001$) and at 20 ms ($p < .05$). For DR, there were statistically significant differences between 7S and 3S sequences at 10 ms ($p < .005$) and at 20 ms ($p < .001$).

4. PART 3: EXPERIMENTS 6-7

4.1 Method

Participants. The same four participants were tested in these experiments as in the previous experiments.

Stimuli. All sequences in Experiments 6-7 consisted of 12 tones each. In the default condition (i.e., no temporal cue present), the tones had durations of 200 ms and were separated by 100-ms silent intervals.

In Experiment 6, eight types of pitch sequence were created. Structured sequences with pitch groups of three tones (G3-S sequences) consisted of a pattern of three notes traversing the C-major scale presented four times in succession, each time a step higher (ascending) or lower (descending) along the scale. The unstructured control sequences (G3-US sequences) had the same beginning pitch, the same number of changes in pitch direction, and the same overall pitch movement as their corresponding structured sequences. In addition, the interval size content of the unstructured sequences and the between-group interval sizes were the same as in the corresponding structured sequences.

Structured sequences with pitch groups of four tones (G4-S sequences) each consisted of a pattern of four tones traversing the C-major scale presented three times in succession, each time two steps higher (ascending) or lower (descending) along the scale. The unstructured control sequences (G4-US sequences) were designed in the same way as described above for the G3-US sequences. All structured sequences had a hierarchical pitch structure designed in accordance with the Deutsch and Feroe (1981) model.

The eight types of sequence were presented in each of four temporal configurations, which differed in the value of a temporal cue suggesting a grouping contrary to that suggested by the pitch structure of the structured sequences. In sequences structured in pitch groups of four tones (G4-S sequences) and in the corresponding unstructured control sequences (G4-US sequences), the silent interval following every third tone was increased, creating temporal groups of three tones. In sequences structured in pitch groups of three tones (G3-S sequences) and in the corresponding unstructured control sequences (G3-US sequences), the silent interval following every fourth tone was increased, creating temporal groups of four tones. The inter-tone intervals following every third or fourth tone were increased by values of 0, 15, 30, and 60 ms. The combination of eight types of pitch sequence with four values of the temporal cue resulted in 32 different sequences.

Figure 4: Examples of two sequences used in Experiment 6. The top sequence is a structured ascending sequence with pitch groups of three tones (G3-S). The bottom sequence is the corresponding unstructured control sequence (G3-US).

In Experiment 7, eight types of pitch sequence were created. Structured sequences with pitch groups of three tones (G3-S sequences) consisted of an arpeggiated C-major triad at a higher structural level embellished on a lower structural level by a chromatic lower neighbor tone. Structured sequences with pitch groups of four tones (G4-S sequences) also consisted of an arpeggiated C-major triad at a higher structural level. The notes of the triad were embellished on a lower structural level by two neighbor tones, creating a turn pattern. The unstructured sequences (G3-US and G4-US) did not allow for a parsimonious representation in accordance with the Deutsch and Feroe (1981) model. They had the same starting pitch, the same overall pitch movement, the same within-group interval size content, and the same between-group interval sizes as their corresponding structured sequences. The descending unstructured sequences were transposed retrogrades of the ascending unstructured sequences; that is, the same pattern of melodic intervals occurred in reverse order. Each of these eight types of sequence was combined with each of the four values of temporal cue (0, 15, 30, and 60 ms), resulting in 32 sequences altogether.

Figure 5: Examples of two sequences used in Experiment 7. The top sequence is a structured ascending sequence with pitch groups of three tones (G3-S). The bottom sequence is the corresponding unstructured control sequence (G3-US).

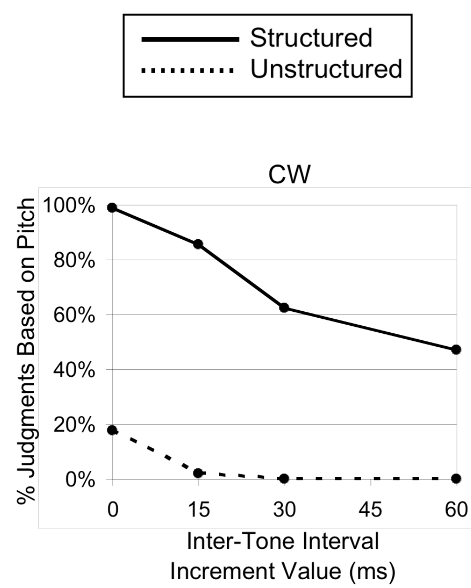
Procedure. In Experiments 6 and 7, each of the 32 different sequences was presented six times over the course of a session in 12 blocks of 16 sequences each. Within each block, the different types of sequence were ordered haphazardly, with the constraints that no more than two sequences with the same value of temporal cue, pitch structure, or pitch group occurred in adjacent positions.

Participants were instructed to judge whether they heard the tones in each sequence as forming groupings consisting of three or of four tones. Each participant completed four experimental sessions over the course of two weeks.

4.2 Results and Discussion

In Experiments 6 and 7, for each of the four participants, data in each pitch structure condition were summed across pitch groups of three and four and ascending and descending sequences, resulting in 96 judgments in each pitch structure/ITI increment condition.

In Experiment 6, structured sequences produced a significantly higher percentage of groupings based on pitch than did unstructured sequences for three of four participants at 0 ms ($p < .001$, for DR, $p > .05$), for all participants at 15 ms ($p < .001$) and 30 ms ($p < .001$; for JK, $p < .05$), and for three of four participants at 60 ms ($p < .001$ for CW and DR; $p < .05$ for RA; $p > .05$ for JK). Figure 6 displays, for each participant, the percentage of groupings based on pitch for structured and unstructured sequences, at each value of temporal cue.



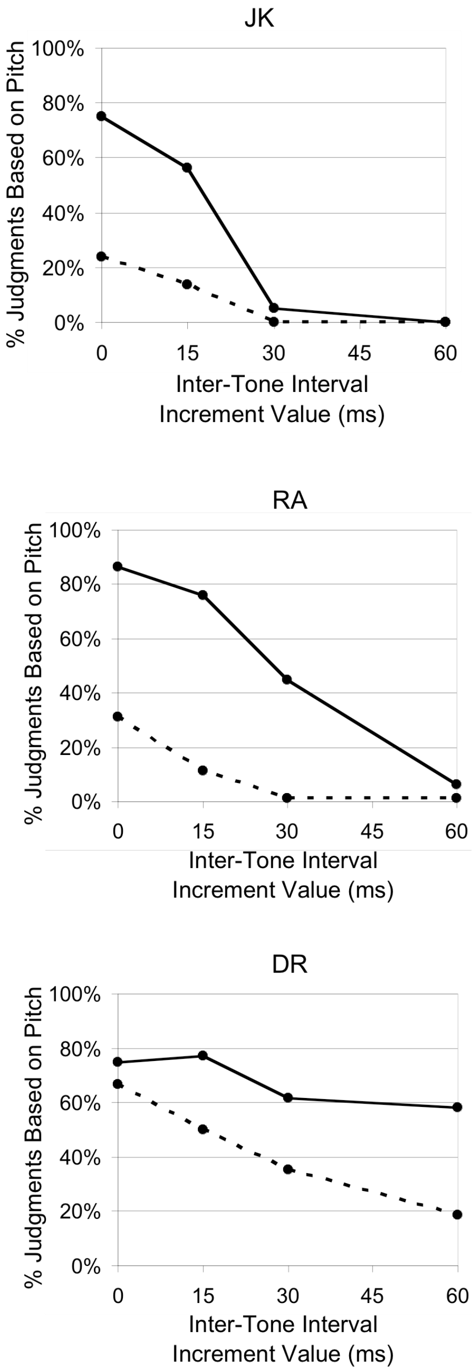
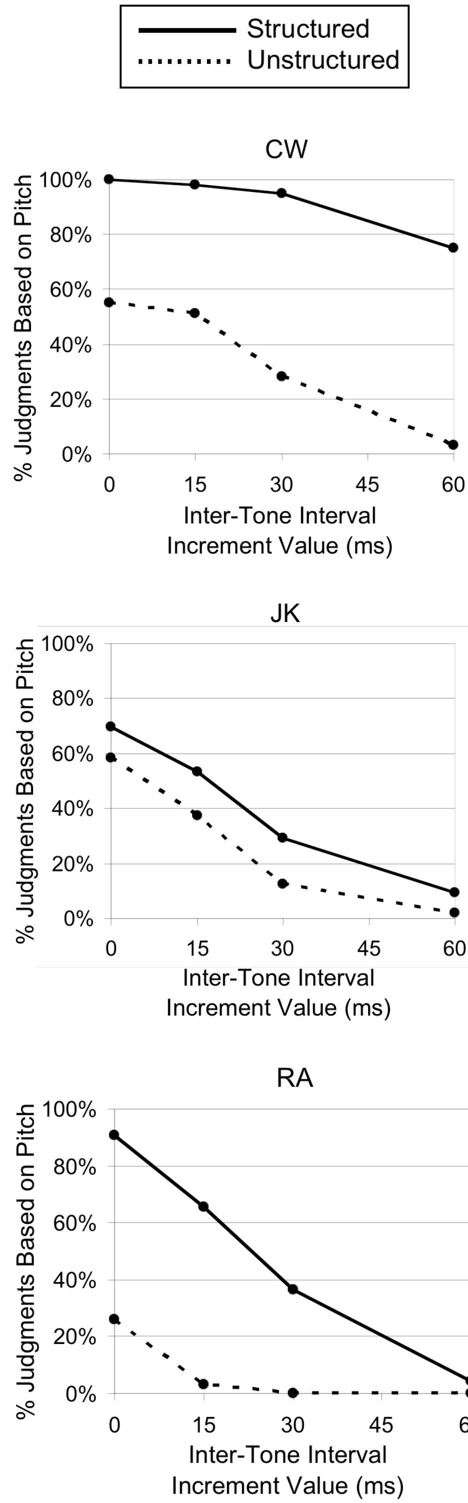


Figure 6. Individual subject data from Experiment 6.

In Experiment 7, structured sequences produced a significantly higher percentage of groupings based on pitch than did unstructured sequences for three of four participants at 0 ms ($p < .001$ for CW and RA; $p < .01$ for DR; $p > .05$ for JK), for all participants at 15 ms ($p < .001$; for JK, $p < .05$) and 30 ms ($p < .001$; for JK, $p < .005$), and for all participants at 60 ms ($p < .001$ for CW and DR; $p < .05$ for JK and RA). Figure 7 displays, for each participant, the percentage of groupings based on pitch for structured and unstructured sequences, at each value of temporal cue.



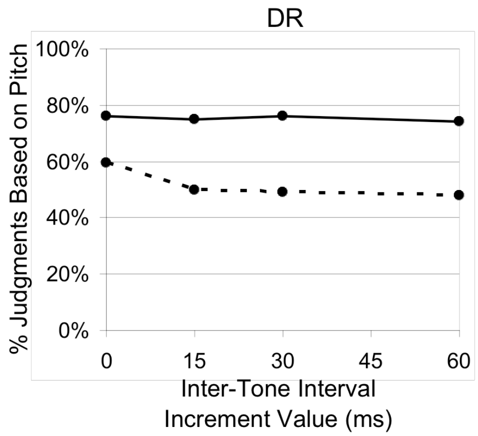


Figure 7. Individual subject data from Experiment 7.

Experiment 6 revealed large effects of hierarchical pitch structure in isolation, when compared with control sequences in which parameters such as interval size and contour were held constant. In fact, for three of the four participants, hierarchical pitch structure had an effect as strong as or stronger than did sequences with eleven-semitone pitch distances between groups.

Experiment 7 demonstrated that hierarchical pitch structure can have additional effects in sequences where pitch proximity is already present as a cue.

5. GENERAL DISCUSSION

The present study presents a new experimental paradigm for studying the perceptual grouping of sound patterns, and demonstrates that this paradigm yields clear and consistent results. In Experiments 1-5, pitch proximity was found to influence perceptual groupings under a variety of experimental manipulations: The larger the pitch distance between tone groups, the stronger the perceptual grouping as indicated by greater resistance to the grouping suggested by the competing temporal cue. The findings of Experiment 6 demonstrate that high-level cues can have an influence on perceptual grouping as great as or greater than the low-level cue of pitch proximity. In addition, the findings of Experiment 7 demonstrate that high-level cues can have an additional effect on perceptual grouping in sequences where the low-level cue of pitch proximity is already present.

In the present study, it was also found that very small values of a temporal cue can influence perceptual grouping. Studies on music perception and performance indicate that performers can use small variations in timing to communicate low-level grouping structure (Sloboda, 1983), and that listeners can make use of these timing cues in their judgments about musical structure (Clarke & Windsor, 2000; Penel & Drake, 2000). The present research suggests that inter-onset interval increments as small as 15-30 ms can influence perceptual grouping for some listeners, though in actual performances we might expect these values to be somewhat larger.

6. REFERENCES

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: MIT Press.

Clarke, E. F., & Windsor, W. L. (2000). Real and simulated expression: A listening study. *Music Perception, 17*, 277-313.

Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl & Jackendoff's grouping preference rules. *Music Perception, 4,* 325-360.

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

Deutsch, D. (1999). Grouping mechanisms in music. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 299-348). San Diego: Academic Press.

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

Dowling, W. J. (1973). Rhythmic groups and subjective chunks in memory for melodies. *Perception & Psychophysics, 4,* 37–40.

Frankland, B. W., & Cohen, A. J. (2004). Parsing of melody: Quantification and testing of the local grouping rules of Lerdahl and Jackendoff's *A Generative Theory of Tonal Music*. *Music Perception, 21,* 499-543.

Handel, S. (1973). Temporal segmentation of repeating auditory patterns. *Journal of Experimental Psychology, 101,* 46–54.

Penel, A., & Drake, C. (2000). Rhythm in music performance and perceived structure. In P. Desain & L. Windsor (Eds.), *Rhythm perception and production* (pp. 225-232). Lisse: Swets & Zeitlinger.

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

Simon, J. L., & Bruce, B. C. (1993). Resampling Stats: User guide (with software) [Computer software and manual]. Arlington, VA: Resampling Stats.

Sloboda, J. A. (1983). The communication of musical metre in piano performance. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 35*, 377-396.

Van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences.* Unpublished doctoral dissertation, Technische Hogeschoel Eindhoven, The Netherlands.