Journal of Experimental Psychology:
Human Perception and Performance
1993, Vol. 19, No. 3, 627–640

0096-1523/93/$3.00

Pitch and Rhythmic Patterns Affecting Infants' Sensitivity to Musical Phrase Structure

Peter W. Jusczyk and Carol L. Krumhansl

The present studies were undertaken to learn more about the nature of the cues that underlie infants' perception of musical phrase structure. Experiment 1 demonstrated that infants in Krumhansl and Jusczyk's (1990) study were responding to the phrase structure instead of to the beginnings and endings of Mozart minuet stimuli. Experiment 2 showed that infants treat musical passages with pauses inserted at phrase boundaries much as they do unaltered versions of the same passages. Experiments 3 and 4 indicated that the direction of change in pitch height and tone duration is critical to obtaining longer orientation times to musical passages that are segmented at phrase boundaries. Finally, Experiment 5 demonstrates that different effects found for the forward and reversed versions of the passages with inserted pauses are not the result of an intrinsic preference for the forward versions.

For many years, psychologists were apt to accept William James's (1890) characterization of the perceptual world of the infant as "one great blooming, buzzing confusion." Even as astute an observer of infants as Piaget (1954) was led to portray young infants as having few resources to deal with the world around them. Basic notions such as space and causality had to be built up from interactions with the environment. Consequently, the infant's world was one in which objects were continually coming into and going out of existence with changes in the infant's own line of sight. However, during the past 20 years, empirical studies of infants have led to a reassessment of their capacities for making sense out of the world around them (for reviews, see Haith & Campos, 1983; Mehler & Fox, 1985; Salapatek & Cohen, 1987). For example, a number of studies have shown that even during the first 6 months of life, infants seem to have some primitive notions of causality (Baillargeon, Graber, DeVos, & Black, 1990; Leslie & Keeble, 1987), number (Antell & Keating, 1983; Starkey, Spelke, & Gelman, 1990), and objects (Kellman, Gleitman, & Spelke, 1987; Spelke, 1985).

Many questions still remain, however, not only about the kinds of information that infants are able to extract from their environments but also about the means by which they accomplish this. Moreover, it is fair to say that the bulk of the research on infants' perceptual capacities has focused on information that is derived through the visual modality (e.g., Haith & Campos, 1983; Salapatek & Cohen, 1987). Yet the developing infant also needs to respond correctly to sources of information that arrive through other sensory modalities such as audition, touch, taste, and smell. Within the auditory realm, a fundamental issue is how infants come to segment complex acoustic signals in a way that identifies the important events, their sources, and the relations among them.

With speech sounds, for example, the listener must be able to analyze utterances into a hierarchy of discrete events such as words, phrases, and clauses. However, as has often been noted (e.g., Hockett, 1954; Klatt, 1979), fluent speech is not segmented into small, discrete packets corresponding to such units. Rather, it is continuous, with one word often running into another. Although fluent speakers of a language may be able to draw on their knowledge of the syntax and semantics of the language to assist in segmenting the input, this avenue is presumably not open to prelinguistic infants. Consequently, infants are forced to rely almost exclusively on potential cues in the acoustic signal as to how the input should be segmented.

Speech is not the only type of acoustic signal in which the issue of segmentation arises. Music is also organized into units of varying temporal extents. A short segment may serve as a unit at one level and then join with other segments to form longer units at higher levels of organization (Deutsch & Feroe, 1981; Lerdahl & Jackendoff, 1983). So just as speech syllables join to form words and words combine into phrases and phrases into clauses, in music tones combine to form melodic and rhythmic figures, musical phrases, and larger sections of musical pieces. Lacking knowledge of any higher level organizational principles that structure music, the infant presumably must rely on cues in the acoustic stream for indications of how it is segmented.

Peter W. Jusczyk, Department of Psychology, State University of New York at Buffalo; Carol L. Krumhansl, Department of Psychology, Cornell University.

This research was supported by National Institute of Child Health and Human Development Research Grant 15795 to Peter W. Jusczyk. We greatly appreciate the assistance of Ann Marie Jusczyk, Tara DeLeon, and Nancy Redanz in testing the infants and Eric Bylund for programming the experiments. J. David Smith, Bruno Repp, and three anonymous reviewers offered numerous useful comments on previous drafts of this article. Finally, we acknowledge the support and encouragement we received for this project from the late Fred Attneave. His questions and comments inspired much of the work reported here.

Correspondence concerning this article should be addressed to Peter W. Jusczyk, Perceptual Development Laboratory, Department of Psychology, 373 Park Hall, State University of New York at Buffalo, Buffalo, New York 14260.

What capacities do infants have for segmenting complex acoustic signals such as speech and music, and on what sources of information do they rely for this purpose? In the case of speech, it is known that boundaries between important grammatical units are often correlated with changes in prosodic variables such as intonation contour, syllable duration, and amplitude level (e.g., Cooper & Sorensen, 1977; Crystal & House, 1988; Klatt, 1975; Lehiste, Olive, & Streeter, 1976; Scott, 1982; Streeter, 1978). Moreover, perceptual studies indicate that the presence of such prosodic markers helps to determine adults' interpretations of syntactically ambiguous sentences (e.g., Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Scott & Cutler, 1984). Recent studies undertaken with infants also suggest that attention to these sorts of prosodic markers may assist infants in segmenting speech. For example, a number of investigators reported that prosodic marking of important syntactic units is not only present but often exaggerated in child-directed speech (e.g., Fernald & Mazzie, 1991; Fisher, 1991; Jusczyk et al., 1992; Lederer & Kelly, 1991; Morgan, 1986). Most important, a growing body of evidence suggests that infants are sensitive to the presence of such prosodic marking in the speech stream. For instance, Hirsh-Pasek et al. (1987) found that 6-month-olds listened significantly longer to speech passages interrupted at clause boundaries than passages interrupted within clauses, a result that was later replicated with 4½-month-olds (see Jusczyk, 1989). More recently, Jusczyk et al. (1992) examined American infants' sensitivity to prosodic markers within English clauses, such as subject and predicate phrases. They found that infants begin to display significant preferences for uninterrupted phrasal units at about 9 months of age.

Studies of segmentation of music suggest that similar processes may underlie its segmentation into underlying structural units. Experiments with adult listeners indicate that the segmentation of music depends on a number of factors including contrasts of pitch range, dynamics, and timbre; lengthening of durations; changes of melodic contour; and metrical, tonal, and harmonic stress (Clarke & Krumhansl, 1990; Deliege, 1987, 1989; Palmer & Krumhansl, 1987a, 1987b; Simon & Sumner, 1968). Furthermore, studies with infants provide considerable evidence that infants are sensitive to a wide range of musical variables, such as changes in melodic contour, rhythm, pitch range, timbre, and dynamics (Chang & Trehub, 1977; Demany, 1982; Demany & Armand, 1984; Thorpe & Trehub, 1989; Thorpe, Trehub, Morrongiello, & Bull, 1989; Trehub, 1987). Nevertheless, it is only recently that the issue of whether infants are sensitive to the segmental structure of music has been addressed directly.

Krumhansl and Jusczyk (1990) investigated whether infants might show preferences for segmentations of musical passages consistent with its underlying structural organization as opposed to segmentations inconsistent with it. If found, this suggests that the infants are sensitive to some of the characteristics of the structural organization of the passages. The experiment used musical passages with relatively obvious and regular melodic, harmonic, and rhythmic structure. The musical passages were all excerpts from simple minuets written by Mozart as a child. Each of the minuets contained an initial section of either 8 or 10 measures followed by a repeat sign. The test stimuli for the experiment were prepared by inserting two-beat pauses (approximately 1 s) in the passages either at the end of each musical phrase (the naturally segmented versions) or by inserting an equal number of pauses in the middle of musical phrases (the unnaturally segmented versions).

The different versions of the samples were presented to infants using a variant of the head-turn preference procedure (see Fernald, 1985; Hirsh-Pasek et al., 1987). Preferences for the alternative versions of the minuets were indexed by measuring the amount of time that infants oriented to the loudspeaker during the playing of each type of sample. In the first experiment, 24 six-month-old infants were tested. The infants had significantly longer orientation times for the naturally segmented versions than for the unnaturally segmented versions of the samples. Twenty-two of the 24 infants showed this pattern, and it held for all 12 minuets tested. A second experiment was undertaken with a comparable group of 4½-month-olds, and the same basic pattern of results was obtained. The mean orientation times for the two versions are shown in Table 1 for the two age groups. Thus, at both ages, the infants displayed strong preferences for the versions of the musical passages in which pauses were inserted at phrase

Table 1
Summary of Experimental Conditions and Average Orientation Times (in Seconds)

	Version						
Study/results	A	B	C	D	E	F	G
Krumhansl and Jusczyk (1990)							
6-month-old infants		10.23 (.63)	8.03 (.46)****				
4½-month-old infants		12.81 (1.08)	10.05 (.80)***				
Present study							
Experiment 1		12.10 (.90)		8.92 (.86)**			
Experiment 2	9.76 (.90)	9.99 (1.08)*					
Experiment 3		11.91 (.85)			8.51 (.71)**		
Experiment 4					8.29 (.87)	9.48 (1.11)*	
Experiment 5	9.85 (.72)						9.28 (1.05)*

Note. A = original music; B = naturally segmented; C = unnaturally segmented (shifted); D = unnaturally segmented; E = reversed naturally segmented; F = reversed unnaturally segmented (shifted); G = reversed original music. Standard errors are in parentheses.
* $p > .05$. ** $p < .001$. *** $p < .005$. **** $p < .0001$.

endings as opposed to in the middle of phrases.

To gain some indication of the musical attributes to which infants were responding, Krumhansl and Jusczyk (1990) examined the musical scores for properties that could potentially correlate with the orientation time differences between the naturally segmented and unnaturally segmented versions of the minuets. One of the variables identified had to do with the pitch of the tones surrounding the pause locations. The top of Figure 1 shows the average pitch of the last three tones (including the bass line) before the pauses and the first two tones after the pauses. Note that there is a sharp drop in the average pitch height before the pauses for the naturally segmented versions of the samples but not for the unnaturally segmented versions. Across both age groups, the pitch height for the last tone before the pause correlated significantly with the orientation times, suggesting that one of the cues that signals the locations of phrase endings is a drop in pitch height. Another of the musical variables identified had to do with the duration of the tones in the melodic line in the vicinity of the pause locations. The bottom of Figure 1 shows the average durations of the last three tones in the melodic line before the pauses and the first two tones after the pauses. Note that there is a sharp increase in the durations of the tones before the pauses for the naturally segmented, but not for the unnaturally segmented, versions of the samples. Once again, there were significant correlations for both age groups between the duration of the last tone before the pause and the mean orientation times, suggesting that a slowing down of the tones in the melodic line may play a role in signaling the phrase endings. A third variable, the identity of the harmonic interval between the last pitch of the melody and its accompanying bass tone (whether or not it was an octave), also correlated significantly with the orientation times for the 6-month-olds. Although in the same direction, the relation was not significant for the 4½-month-olds.

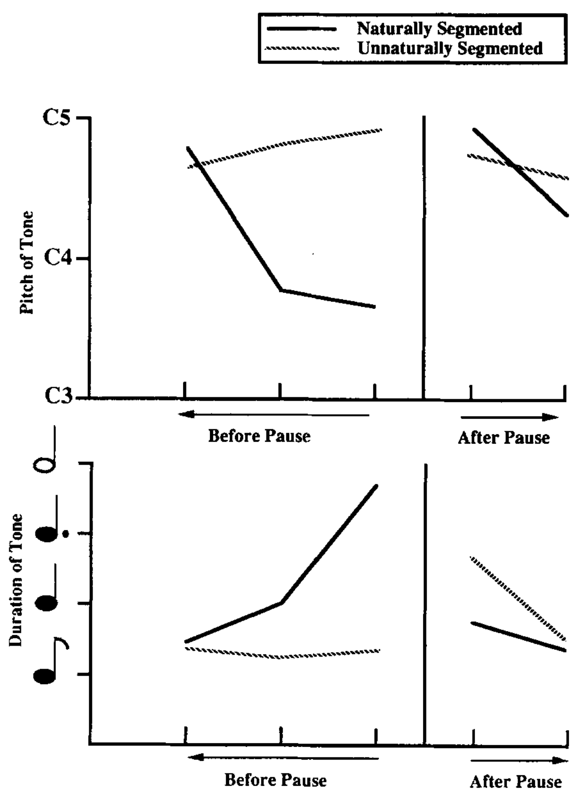

Figure 1. Analyses of properties of the musical samples that correlated significantly with infants' orientation times in the study by Krumhansl and Jusczyk (1990). (Top: the average pitch of the last three tones before the pauses and the first two tones after the pauses for the naturally segmented [solid line] and the unnaturally segmented [broken line] versions. Bottom: the average duration of the last three tones in the melody line before the pauses and the first two after the pauses for the naturally segmented [solid line] and the unnaturally segmented [broken line] versions.) Reprinted from "Infants' perception of phrase structure in music," by Krumhansl & Jusczyk, 1990, *Psychological Science*, 1, Figure 2. Copyright Cambridge University Press. Reprinted by permission.

It is interesting to note that two of the musical variables that correlated highly with the infants' orientation times appear to parallel the acoustic cues that Jusczyk et al. (1992) suggested might underlie infants' segmentation of speech samples. In particular, the pitch height cue that signals phrase endings in the musical passages is similar to the change in fundamental frequency that occurs at the ends of clauses and phrases, and the change in tone duration at the ends of musical phrases parallels the lengthening of final syllables before important syntactic boundaries. The similarities observed across these two domains raise the possibility that infants may be responding to general properties of acoustic signals that serve to mark important events during auditory perception.

The present studies were undertaken in an effort to learn more about the nature of the cues that may underlie infants' segmentation of musical passages. In particular, we focused on the variables that Krumhansl and Jusczyk (1990) identified as correlating with the longer orientation times that infants displayed for the naturally segmented passages. As a first step, it was necessary to examine some potential alternative explanations for the pattern of results that Krumhansl and Jusczyk observed. Experiment 1 investigates one such possibility having to do with the beginnings and endings of the naturally segmented and unnaturally segmented versions of the minuets. Experiment 2 explores the possibility that the very act of inserting pauses into the samples may have resulted in stimuli that are unnatural regardless of whether the pauses occur at the phrase endings. Experiment 3 presents a test of whether the direction of change in the pitch height and pitch duration cues are critical to obtaining longer orientation times to musical passages that are segmented at phrase endings. This is done by presenting the musical passages with the tones played in both forward and reversed orders and segmenting them at phrase boundaries. The hypothesis that the direction of the pitch height and pitch duration changes are important receives a further test in Experiment 4 in which infants are exposed to reversed versions of the same passages that are segmented either at phrase endings or in the middle of phrases. Finally, Experiment 5 examines whether differences between the way in which infants respond to forward and reversed passages reflects a simple preference for the forward over the reversed passages.

Experiment 1

A number of factors must be considered when one prepares naturally segmented and unnaturally segmented versions of the same musical passage. One can start both versions at the beginning of a piece and simply place an equal number of pauses after different measures in the two versions of the same musical sample. This has the consequence that the absolute location of a pause (e.g., whether it occurs at the end of the second measure or some other measure) will differ for the naturally and unnaturally segmented versions. Alternatively, one can attempt to maintain the same absolute location of the pauses (e.g., after the second measure for both versions) while varying where in the passage the unnaturally segmented version starts relative to the naturally segmented version. Krumhansl and Jusczyk (1990) chose the latter approach by beginning the music after one, three, five, seven, or nine measures. To illustrate, Figure 2 shows one of the stimulus passages, which consists of phrases of lengths two, two, four, two, two, and four measures. This phrase structure is shown schematically at the top of Figure 3 (Version A: original music). The naturally segmented version (Version B) was created by inserting two-beat pauses between phrases. The unnaturally segmented (shifted) version (Version C) was created by beginning the music in Measure 2 and inserting the pauses in the same absolute locations as the naturally segmented version. This approach to creating unnaturally segmented versions (by shifting the music) controls for the possibility that differences in orientation times might have nothing to do with how pauses are positioned with respect to the phrase structure of the passages but might instead reflect a preference for the absolute location of the pauses (e.g., a pause after two measures might be more pleasing than one after three measures or vice versa).

Although this approach allowed Krumhansl and Jusczyk (1990) to ensure that longer orientation times to the naturally segmented versions did not simply reflect the absolute location of the pauses, it does not rule out another alternative explanation of the results. This alternative has to do with the possibility that the infants were responding to the passages on the basis of the way in which they began instead of where the pauses were located with respect to the phrase structure of the passages. That is, the earlier result may have depended on the fact that the naturally segmented versions always began at the beginning, whereas the unnaturally segmented (shifted) versions began at some other point in the music. To investigate this potential alternative explanation of the results, we decided to attempt to replicate the Krumhansl and Jusczyk (1990) findings using materials in which the naturally segmented and unnaturally segmented versions always began at the beginning of the music. The locations of the pauses in the naturally segmented versions were the same as in the earlier study. The locations of the pauses in the unnaturally segmented versions followed the same pattern but were shifted by one, three, five, seven, or nine measures. This was done so that, relative to the music, the pauses occurred in the same location in the unnaturally segmented versions in this and the earlier experiments. Figure 3 illustrates the two conditions used in Experiment 1 as Version B, naturally segmented, and Version D, unnaturally segmented. Because Krumhansl and Jusczyk (1990) found essentially the same pattern of results for both 4½- and 6-month-olds, we decided to focus our investigation only on the younger age group.

Method

Subjects. Subjects were twenty-four infants approximately 4½ months of age (mean age = 19.8 weeks; range = 19.0–21.3 weeks); 2 additional infants were tested but did not complete the experimental session.

Apparatus. The stimulus materials were produced by a Yamaha TX816 FM tone generator synthesizing a piano timbre under the control of a Macintosh Plus computer using the Performer software. The stimuli were recorded on a Revox reel-to-reel tape recorder and then digitized by a PDP 11/73 computer using a 12-bit A/D converter. During the experiment, the computer controlled the presentation of the trials and recorded the observer's coding of the infants' responses. The audio output for the experiment was generated from the digitized waveforms of the samples stored on the computer. A 12-bit D/A converter fed the output through antialiasing filters and a Kenwood audio amplifier (Model #KA 5700) to the 17.5-cm (7-in.) loudspeakers mounted on the sidewalls of the test booth.

The experiment was conducted in a test booth constructed out of pegboard, with panels of 120 cm × 180 cm (4 ft × 6 ft) on three sides and open at the back. This made it possible for an observer to look through one of the existing holes in the pegboard to monitor the infant's head turns. Except for a small section for viewing the infant, the remainder of the pegboard was backed with white cardboard to guard against the possibility that the infant might respond to movements behind the panel. The test booth had a red light and a loudspeaker mounted at the seated infant's eye level on each of the side panels and a green light mounted on the center panel. A white curtain suspended around the top of the booth shielded the infant's view of the rest of the room. A computer terminal and response box were located behind the center panel out of view of the infant. The response box,

Figure 2. One of the excerpts from a minuet used in the experiments. (It consists of eight measures followed by a repeat sign giving a total of 16 measures. It is composed of phrases of lengths two, two, four, two, two, and four measures.)

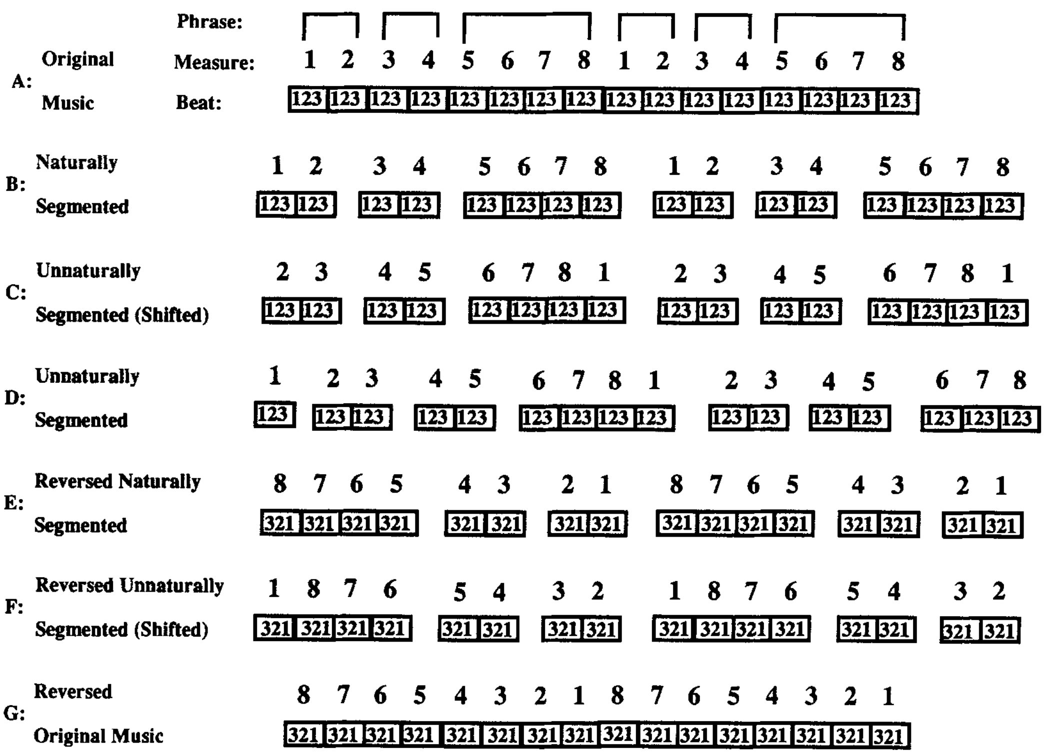

Figure 3. Schematic demonstrating the way that pauses were inserted into the music to create the various versions used in the experiments. (For purposes of illustration, the phrase structure is that of the excerpt shown in Figure 2. Version A: Original music is the music as written, played without pauses. Version B: Naturally segmented is the music played with pauses inserted between phrases. Version C: Unnaturally segmented [shifted] is the music shifted [by one measure in the case illustrated] so that the pauses occur in the middle of phrases. Version D: Unnaturally segmented is the music played from beginning to end with the pauses shifted so that they occur in the middle of phrases. Version E: Reversed naturally segmented is the same as Version B, except the tones are played in reversed order. Version F: Reversed unnaturally segmented [shifted] is the same as Version C, except the tones are played in reverse order. Version G: Reversed original music is the same as Version A, except the tones are played in reversed order. Table 1 shows the conditions used in each of the experiments.)

which was connected to the computer, was equipped with a series of buttons that started and stopped the flashing center and sidelights, recorded the direction and duration of head turns, and terminated a trial when the infant looked away for more than 2 s. The experimenter pushed one button to initiate a trial, a second button to indicate that the infant was turned in the direction of the flashing sidelight, and a third button to indicate that the infant was looking away from the sidelight. Information about the direction and duration of head turns and the total trial duration was stored in a data file on the computer.

Stimulus materials. Sixteen Mozart minuets were selected from *Trente Pièces Faciles Pour Piano* (Thirty Easy Pieces for Piano, Schott Frères, Brussels). Each piece contained an initial section of 8 or 10 measures followed by a repeat sign. These 16 or 20 measure-long sections (the initial section and its repetition) were used to construct the naturally segmented and unnaturally segmented versions. In the naturally segmented versions, a two-beat pause (duration = .96 s) was inserted at the end of each phrase. The phrase locations were determined by the experimenters' intuitions (and were confirmed by adult listeners' judgments; Krumhansl & Jusczyk, 1990). Each stimulus consisted of either 4 or 6 phrases (of measure lengths 2, 4, or 6). The unnaturally segmented versions contained the same pattern of pauses only shifted relative to the music (by 1, 3, 5, 7, or 9 measures) so that the pauses did not correspond to natural phrase endings. All the minuets were notated in three-quarter time and played at a tempo of 125 quarter notes per minute.

Procedure. The procedure was a modified version of one originally developed by Fernald (1985). Each infant was held on a parent's lap. The parent was seated in a chair in the center of the test booth. The infant completed an 8-trial familiarization phase (both versions of four minuets) and a 12-trial test phase. The naturally segmented versions were consistently played through the loudspeaker on one side panel, and the unnaturally segmented versions through the loudspeaker on the other side panel. (The side was counterbalanced across subjects.) The familiarization phase was intended to acquaint the infant with the assigned position of each type of version. Minuet numbers 1, 3, 4, and 5 were used during the familiarization period. For the 12 test trials that followed, half of the infants heard the naturally segmented and unnaturally segmented versions of six of the minuets (numbers 7, 8, 9, 12, 13, and 14), and the other half heard the two versions of the other six minuets (15, 16, 17, 18, 23, 24). The ordering of the stimuli was random subject to the constraint that, for three of the minuets, the naturally segmented version occurred before the unnaturally segmented version and vice versa.

Each trial began by blinking the green light on the center panel until the infant had oriented in that direction. Then the experimenter pressed a button that caused the center light to be extinguished and, at the same time, caused the red light above the loudspeaker on one of the side panels to begin to flash. When the infant made a head turn of approximately 30° in the direction of the loudspeaker and flashing red light, the experimenter pressed a second button. This started the musical sample. It continued to play until its completion or until the timing of the button presses by the experimenter indicated that the infant failed to maintain the 30° head turn (e.g., if the infant turned back to the center or the other side, looked at the mother, the floor, or the ceiling). If 2 consecutive s elapsed without the experimenter pressing the button to indicate that the infant was now once again oriented to the flashing red light, the computer terminated the trial. If the infant turned briefly away from the target by 30° in any direction, but for less than 2 s, and then looked back again, the time spent looking away was not included in the orientation time and the sample continued to play. During the familiarization trials, the red light was extinguished when the music began; however, during the test trials the light remained on for the entire duration of the trial.[1]

An observer hidden behind the center panel looked through a peephole and recorded the direction and duration of the infant's head turns using a response box. The observer was not informed as to which loudspeakers played the naturally segmented and unnaturally segmented versions. This was possible because the assignment of the versions to the left or right side was determined by the computer and not revealed to the observer until the completion of the test session. The loudness levels for the samples were set before each test session by a second assistant, who was not involved in the observations, at 72 ± 2 dB (C) SPL using a Quest (Model 215) sound level meter. In addition, both the observer and the infant's parent listened to recorded music over headphones so they were unaware of the location or the nature of the stimulus on the trial. Testing took between 5 and 7 mins.

Results and Discussion

The amount of time that the infant oriented to the loudspeaker playing each sample was measured. The mean orientation times, shown in Table 1, were longer for the naturally segmented versions than for the unnaturally segmented versions. Twenty-one of the 24 infants showed this pattern, and the difference was highly statistically significant, $t(23) = 4.19$, $p < .001$. Moreover, for 11 of the 12 minuets, the mean orientation time was longer for the naturally segmented version than for the unnaturally segmented version, $t(11) = 3.635$, $p < .005$.

The musical variables that correlated with the orientation times in the earlier study (Krumhansl & Jusczyk, 1990) were correlated with the present data. The first analysis considered the pitch of the last tones (including the bass) before the pauses. (When two tones were sounded simultaneously, the tone with the higher pitch was included in the analysis.) The pitch of the last tones before the pauses was lower on average in the naturally segmented versions than in the unnaturally segmented versions. If middle C (C_4) is coded as 60, the average pitch of the last tones before the pauses was 56.60 (i.e., between $G\#_3$ and A_3) in the naturally segmented versions and 72.25 (i.e., between C_5 and $C\#_5$) in the unnaturally segmented versions. The analysis focused on the pitch of the last tone because this variable correlated significantly with the orientation times in the earlier study. It should be noted that the lower last pitches in naturally segmented versions mean these exhibit greater drops in pitch before the pauses than the unnaturally segmented versions (see the top of Figure 1). In this experiment, the average pitch of the last tones before the pauses in each sample correlated significantly with orientation time, $r(22) = -.39$, $p < .05$ (one-tailed test); as before, longer orientation times were associated with lower ending pitches.

The second analysis considered the duration of the last melody tones before the pauses. The average duration of the last tones was 1.92 beats (.92 s) in naturally segmented versions and .66 beats (.32 s) in unnaturally segmented versions. This variable was chosen for the analysis because it correlated significantly with the orientation times in the earlier study. The longer durations of the last tones for the naturally segmented versions mean that there is a greater increase in durations before the pauses in these versions than in the unnaturally segmented versions (see the bottom of Figure 1). The average duration of the last tones before the pauses in each sample correlated significantly with orientation times, $r(22) = .37$, $p < .05$ (one-tailed test); as before, longer orientation times were associated with longer durations of the melody tones before pauses.

The final analysis considered the interval formed by the last tone of the melody before the pauses and its accompanying bass tones. The proportion of intervals preceding pauses that were octaves was higher in naturally segmented versions (.56) than in unnaturally segmented versions (.11). The proportion of octave intervals before pauses in each sample correlated significantly with orientation times, $r(22) = .35$, $p < .05$ (one-tailed test); as before, longer orientation times were associated with octave intervals in the final positions before the pauses.[2]

The results of the present study fully replicate those of Krumhansl and Jusczyk (1990) with respect to both the

[1] Extensive pilot testing before previous studies (e.g., Jusczyk et al., 1992) indicated that this was the best method for operating the sidelights during the procedure. Leaving the flashing light on during the familiarization trials seems to habituate the infants to the lights and results in very short looking times during the test trials. Moreover, infants are less likely to complete the full set of test trials under these circumstances.

[2] The three variables considered in these analyses (pitch height, duration, and proportion of octave intervals) were among a number of variables that were considered in connection with the data of the Krumhansl and Jusczyk (1990) study. These included the function of tones surrounding pauses, contour direction, the consonance of the final interval preceding the pause, and the size of the interval formed by tones immediately before and after the pauses. None of these variables correlated as strongly with the orientation time measures as the three variables considered here. Because the naturally segmented and unnaturally segmented versions differ markedly in terms of these three variables, it is possible that they are redundant with the distinction between naturally segmented and unnaturally segmented versions. However, additional analyses showed the combined effect of the three variables is stronger than that of a single variable coding the two types of stimuli.

longer orientation times displayed for the naturally segmented versions of the minuets, and the correlation of these orientation times with the musical variables of pitch height and tone duration. The third musical variable, the proportion of octave intervals preceding pauses, also correlated significantly with the 4½-month-olds' orientation times in the present study; in the earlier Krumhansl and Jusczyk (1990) study, a significant correlation was found only for the 6-month-old subjects. The present replication suggests that the infants in the earlier study were not simply responding to the way that the naturally segmented and unnaturally segmented versions of the samples began. Instead, it appears that the primary factor affecting infants' orientation times has to do with where the pauses are located with respect to the phrase structure of the passages. However, one might still object that the introduction of the pauses somehow created an artificial or "super" stimulus, and that, without the inserted pauses, infants are not sensitive to the musical phrase structure of these pieces. To explore this possibility, we decided to conduct the following experiment.

Experiment 2

One possible consequence of introducing pauses into our samples is that it greatly increases the salience of the phrase structure of the piece. The presence of the pauses may provide the infant with additional information about the structure of the pieces that is not available without the inserted pauses. In this connection, it should be noted that performers often add expressive timing variations (e.g., slowing tempo approaching a phrase boundary and pauses between phrases) to emphasize the musical structure. Perhaps the additional information of the inserted pauses is necessary for the infants to discover the phrase structure of the pieces. It may be, for example, that the placement of pauses draws the infants' attention to musical features at certain points in the music and thus leads them to prefer the naturally segmented versions to the unnaturally segmented versions. If this view is correct, then one might expect to find that infants actually prefer to listen to the naturally segmented versions (with pauses) more than the unsegmented versions (without pauses) of the minuets. Alternatively, it is possible that adding pauses to the samples renders them somehow artificial for the infant. If so, then they might actually prefer to listen to the unsegmented versions of the minuets more than the naturally segmented versions. To gain a better understanding of the perceptual consequences of inserting pauses in the samples, we decided to examine how 4½-month-olds respond to the original music versions (Version A; see Figure 3) as opposed to the naturally segmented versions (Version B; see Figure 3) of the same minuets.

Method

Subjects. Subjects were 24 infants approximately 4½ months of age (mean age = 21.0 weeks; range = 19.0–22.8 weeks); 4 additional infants were tested but did not complete the experimental session.

Apparatus. The apparatus was the same as in Experiment 1.

Stimulus materials. The stimuli were of two types: original music and naturally segmented. The original music versions simply consisted of the original music without added pauses. The naturally segmented versions were the same as in Experiment 1 (the original music with added pauses).

Procedure. The procedure was the same as in Experiment 1, except that minuet numbers 1, 3, 7, and 24 were used during the familiarization trials, and one half of the infants heard the two versions of minuet numbers 4, 5, 8, 14, 15, and 16, whereas the other half heard the two versions of minuet numbers 9, 12, 13, 17, 18, and 23.

Results and Discussion

The mean orientation times, shown in Table 1, showed no difference between the original music and the naturally segmented versions. Eleven of the 24 infants preferred the original music versions to the naturally segmented versions, and the difference was not statistically significant, $t(23) = -.20$, $p > .05$. For 6 of the 12 minuets, the original music version was preferred to the naturally segmented version, and the difference was not statistically significant, $t(11) = -.30, p > .05$. Thus, there was no indication that infants preferred listening to one version of a sample more than to the other. Consequently, it seems unlikely that they treat the naturally segmented versions of the samples either as some kind of super stimulus or as more artificial than the original music versions of the minuets.

Although the present study cannot completely rule out the possibility that infants still perceive the original music and naturally segmented versions of the stimuli differently, it helps allay the concern that inserting pauses in the pieces significantly affects how infants respond to them. Thus, we can now return to the question of the nature of the cues that may underlie infants' perception of phrase structure in these pieces.

Experiment 3

In the studies that we have conducted comparing infants' orientation times to naturally segmented and unnaturally segmented versions of the same musical pieces, we have noted certain musical variables in the scores of these pieces that correlate significantly with the longer orientation times to the naturally segmented versions. In particular, pitch height tends to fall and tone duration tends to increase at musical phrase endings. Also, a greater proportion of octave intervals occurs at phrase endings in the naturally segmented versions than in the unnaturally segmented versions. As noted earlier, the first two of these variables, pitch height and tone duration, appear to be analogous to the fundamental frequency and syllable duration changes that have been implicated as cues for segmenting speech into clauses and phrases (e.g., Cooper & Sorensen, 1977; Crystal & House, 1988; Klatt, 1975; Lehiste et al., 1976; Price et al., 1991; Scott, 1982; Streeter, 1978). It is tempting to speculate that perceptual segmentation of successive auditory events might rely on the presence of certain acoustic markers such as falls in pitch and increased durations of elementary components (e.g., syllables in the case of speech, tones in the case of music). Thus, the presence of such markers could be an indication that an event in progress is coming to an end. However, although our analyses of the musical scores show significant correlations between the presence of such variables and the infants' orientation times, we cannot completely rule out the possibility that other factors may be responsible for the infants' preferences for the naturally segmented over the unnaturally segmented versions of the samples.

Reconsidering the musical scores in the vicinity of the phrase boundaries reveals another way of characterizing the factors that may be determining the infants' preferences for the naturally segmented versions. There are marked changes in both pitch height and tone durations across the phrase boundaries.[3] Consider the pitch height variable (as illustrated in Figure 1). For the naturally segmented versions, the melodic line drops sharply before the pause, and the pitch height of the tones immediately after the pause is considerably higher. In contrast, for the unnaturally segmented versions, no such discontinuity is present between the tones immediately before the pause and the tones immediately after it. Could it be that it is the presence of the discontinuity, and not the pitch fall per se, in the naturally segmented versions of the minuets that serves to indicate the phrase boundary? Similarly, a discontinuity is also present in the case of the tone duration variable for the naturally segmented versions, but not for the unnaturally segmented versions, of the minuets. Specifically, for the naturally segmented versions, the tone durations increase immediately before the pause, and the tone durations immediately after the pause are much shorter. Once again, for the unnaturally segmented versions, there is much more continuity between tone durations before and after the pause. This raises the question: Is it merely the presence of a discontinuity with respect to the pitch height and tone duration variables that is important, or does the direction of the change also matter?

One way of investigating this issue is to use samples that contain the same magnitude of discrepancy before and after the pauses in pitch height and tone duration but that differ in the direction of the changes. There are a number of ways that one might attempt to keep the discontinuity while reversing the directionality but, unfortunately, most of these involve rewriting the music with consequent changes in such factors as pitch content, complexity of melodic contour, and rhythm. The new versions may simply be less coherent or aesthetically pleasing than the old ones, making it difficult to interpret any failure to find a difference between the two versions. After considering a number of possibilities, we decided to simply play the minuets with the tones in the reversed order, keeping the durations between tone onsets the same as before. The pauses still occur between the same groups of tones, but now the relative pitch heights and tone durations in the vicinity of the phrase boundaries are reversed. The forward and reversed versions have a number of important characteristics in common. They contain exactly the same pitches and harmonic intervals and the same complexity of melodic contour. Subjectively, they have similar tonal orientations (owing to the simple tonal structure of the pieces) and clearly have the same meter. By comparing the naturally segmented versions of the minuets with the same music played in the reversed direction, we can gain some indication of whether the directionality of the change in pitch and duration is important. So, in this experiment, we pitted the naturally segmented versions (Version B; see Figure 3) against the reversed versions (the same minuets played beginning with the last note of the score and ending with the first note; Version E; see Figure 3). If the directionality of the change in musical variables is important, then we expect infants to show longer orientation times to the naturally segmented versions.

Method

Subjects. Subjects were 24 infants approximately 4½ months of age (mean age = 20.3 weeks; range = 19.1–21.7 weeks); 2 additional infants were tested but did not complete the experimental session.

Apparatus. The apparatus was the same as in Experiment 1.

Stimulus materials. The stimuli were of two types: naturally segmented and reversed naturally segmented. The naturally segmented versions were the same as those used in Experiment 1. The reversed naturally segmented versions were the same sequences but played with the tones in reversed order. The transformation was effected by reversing the order of the tones in the computer score, which was then played by the synthesizer.

Procedure. The procedure was the same as in previous experiments, with the same assignment of minuets to familiarization and test trials as in Experiment 2.

Results and Discussion

The mean orientation times for the two conditions, shown in Table 1, show a preference for the naturally segmented versions over the reversed naturally segmented versions. Eighteen of the 24 infants preferred the naturally segmented versions to the reversed naturally segmented versions, and the difference was statistically significant, $t(23) = 4.10, p < .001$. For all of the 12 minuets, the naturally segmented version was preferred to the reversed naturally segmented version, and the difference was statistically significant, $t(11) = 4.213, p < .01$. Thus, the naturally segmented versions were quite consistently preferred to the reversed naturally segmented versions.

The average pitch of the last tones before the pauses was 56.07 (i.e., between $G\#_3$ and A_3) in the naturally segmented versions and 69.17 (i.e., between A_4 and $A\#_4$) in the reversed naturally segmented versions. This variable correlated strongly with orientation times, $r(22) = -.62, p < .001$ (one-tailed test), with longer orientation times associated with lower ending pitches. The durations of the melody tones before the pauses averaged 1.83 beats (.88 s) in the naturally segmented versions and were always equal to 1.0 beats (.48 s) in the reversed naturally segmented versions. (This is because in the original music the first tones after the pauses were always on the first beat of the measure so that, when the music was reversed, the last tones before the pauses were

[3] The late Fred Attneave first drew our attention to this point.

always on the last beat of the measure.) The durations of the tones before the pauses correlated significantly with orientation times, $r(22) = .45$, $p < .05$ (one-tailed test), with longer orientation times associated with longer durations of the melody tones before pauses. The proportion of intervals before the pauses that were octaves was quite similar in the naturally segmented (.64) and reversed naturally segmented (.58) versions. This variable did not correlate significantly with orientation times, $r(22) = .03$, $p > .05$.

The present results support the view that the directionality of pitch height and tone duration had a significant impact on infants' orientation times to the samples. Thus, even though the pauses for the naturally segmented and reversed naturally segmented versions of the minuets were inserted between the same sequences of tones, the infants listened significantly longer to the sequences when falls in pitch and increases in tone duration preceded the inserted pauses. Note that there were discontinuities in the pitch height and tone duration variables across the pauses for both the naturally segmented and reversed naturally segmented versions of the minuets. This suggests that the direction of changes in the musical variables before the pause is more important than the simple presence of a discontinuity in these variables. The present results, however, do not completely rule out the possibility that infants might be responding to such discontinuities in addition. This is because discontinuities occurred in both versions of the minuets in the present experiment. By comparison, the naturally segmented and unnaturally segmented versions of the minuets used in Experiment 1 (and in Krumhansl & Jusczyk, 1990) differed both in the direction of the pitch change and tone duration variables and in whether a discontinuity occurred in these variables at the points where the pauses were inserted. So it is possible that the simple presence of discontinuities in pitch height and tone duration, regardless of their directions of change, could also contribute to infants' listening preferences. To examine this possibility, we conducted the following experiment.

Experiment 4

To evaluate whether discontinuity by itself might affect infants' preferences for the musical samples, one must find a setting in which discontinuity is contrasted with no discontinuity. One such setting is suggested by the previous experiment, namely, by playing both the naturally segmented and the unnaturally segmented versions of the minuets with the tones in reversed order. For the reversed naturally segmented versions, the discontinuities in pitch height and tone duration variables before and after the pauses are still present, but the directionalities of the changes are reversed. Thus, instead of a pitch drop just before the pause, there is a large pitch drop after the pause. In contrast, the reversed unnaturally segmented versions have no such discontinuity in pitch height in the vicinity of the pauses. Similar differences hold with respect to the tone duration variable. For the reversed naturally segmented versions, the tone durations immediately after the pause will tend to increase relative to the last tone durations immediately before the pause. This is not the case for the reversed unnaturally segmented versions. Thus, if the presence of discontinuities is important in itself for the infants' perception of musical phrase structure, we can expect significantly longer orientation times to reversed naturally segmented versions (Version E; see Figure 3) of the minuets than to reversed unnaturally segmented (shifted) versions (Version F; see Figure 3).

Method

Subjects. Subjects were 24 infants approximately 4½ months of age (mean age = 20.3 weeks; range = 18.9–22.4 weeks); 4 additional infants were tested but did not complete the experimental session.

Apparatus. The apparatus was the same as in Experiment 1.

Stimulus materials. The stimuli were of two types: reversed naturally segmented and reversed unnaturally segmented (shifted). They were the same sequences used in Krumhansl and Jusczyk (1990) but played with the tones in the reversed order.

Procedure. The procedure was the same as in previous experiments, with the same assignment of minuets to familiarization and test trials as in Experiment 2.

Results and Discussion

The mean orientation times, shown in Table 1, were not different for the reversed naturally segmented and reversed unnaturally segmented versions. Nine of the 24 infants preferred the reversed naturally segmented versions to the reversed unnaturally segmented versions, and the difference was not statistically significant, $t(23) = -1.581$, $p > .05$. For 4 of the 12 minuets, the reversed naturally segmented version was preferred to the reversed unnaturally segmented version, and the difference was not statistically significant, $t(11) = -1.711$, $p > .05$. Thus, there is no evidence that the reversed naturally segmented versions were preferred to the reversed unnaturally segmented versions and, in fact, the difference tended in the opposite direction.

The average pitch of the last tones before the pauses was 69.17 (i.e., between A_4 and $A\#_4$) in the reversed naturally segmented versions and 71.64 (i.e., between B_4 and C_5) in the reversed unnaturally segmented versions. The two versions did not differ consistently in terms of the pitch of tones before the pauses, and this variable did not correlate with orientation times, $r(22) = -.20$, $p > .05$. The durations of the melody tones before the pauses were equal in the two versions. They were always equal to 1.0 beat because in the original music, the first tones after the pauses were always on the first beat of the measure so that, when the music was reversed, the last tones before the pauses were always on the last beat of the measure. The proportion of intervals before the pauses that were octaves was somewhat higher in the reversed naturally segmented versions (.59) than in the reversed unnaturally segmented versions (.32). This variable correlated negatively with orientation times, $r(22) = -.41$, $p < .05$, opposite the effect found in previous experiments. Apparently, this cue may not operate in isolation from the other phrase-signaling cues.

To sum, the results of this experiment indicate that the mere presence of discontinuities in pitch height and tone duration variables in the vicinity of the pauses was not sufficient to induce significantly longer orientation times to the musical samples. Rather, it appears that it is the direction of change in these variables (i.e., a drop in pitch height and an increase in tone duration) that is the critical factor affecting infants' orientation times to the two types of samples. Thus, the pattern of results from the experiments that we have conducted so far suggests that preferences for one type of sample over another emerge only when the types differ as to whether or not they contain drops in pitch height and increases in tone duration before the pauses. Whenever the versions of samples differed in this way, as they did in Krumhansl and Jusczyk (1990) and in Experiments 1 and 3 of the present study, the infants oriented longer to the versions with the falls in pitch height and increases in tone duration preceding the pauses. When the variables were not present in either version (as in Experiment 4), there was no significant difference in the orientation times between the two versions of the samples.

Experiment 5

The pattern of results so far quite strongly argues that drops in pitch and longer tone durations are important cues to phrase structure for the infants. Still one might object that an alternative account could be offered for the results of Experiments 3 and 4. Both experiments used reversed versions of the Mozart minuets. Suppose that infants merely perceived these versions as somehow inferior to the music played in the forward direction. In this case, then, they might prefer to listen to forward versions over reversed versions, as in Experiment 3, and when faced with a choice between two different reversed versions, as in Experiment 4, they may show no preference for either. To rule out the possibility that infants simply preferred the forward to the reversed versions, we conducted an additional experiment. No pauses were inserted into any of the samples. The infants listened to the original music versions (Version A; see Figure 3) and the reversed original music versions (Version G; see Figure 3) of the same minuets. If the reversed versions were inherently less appealing to the infants, then one would expect the infants to have significantly shorter orientation times for the reversed versions than for the forward versions.

Method

Subjects. Subjects were 24 infants approximately 4½ months of age (mean age = 19.4 weeks; range = 17.3–21.6 weeks); 5 additional infants were tested but did not complete the experimental session.

Apparatus. The apparatus was the same as in Experiment 1.

Stimulus materials. The stimuli were of two types: original music and reversed original music. The original music versions were the same as those used in Experiment 2. The reversed original music versions were the same sequence but played with the tones in the reversed order. No pauses were inserted into either version.

Procedure. The procedure was the same as in previous experiments, with the same assignment of minuets to familiarization and test trials as in Experiment 2.

Results and Discussion

The mean orientation times, shown in Table 1, were not different for the original music versions and the reversed original music versions. Fifteen of the 24 infants preferred the original music versions to the reversed original music versions, and the difference was not statistically significant, $t(23) = 0.61, p > .05$. For 7 of the 12 minuets, the original music version was preferred to the reversed original music version, and the difference was not statistically significant, $t(11) = .71, p > .05$. Thus, the original music versions were not consistently preferred to the reversed original music versions.

The implication of these results is that the infants do not simply prefer the forward versions to the same music with the tones played in the reversed order. At first, this may seem surprising, but, as noted earlier, the forward and reversed versions are similar in a number of important respects: pitch content, harmonic intervals, complexity of contour, tonal orientation, and meter. Given the results of this experiment, it seems unlikely that an effect of reversing the order of the tones can account for the preference observed in Experiment 3 for the naturally segmented versions over the reversed naturally segmented versions of the minuets or the lack of preference for the reversed naturally segmented versions over the reversed unnaturally segmented versions in Experiment 4. Rather, the results appear to depend on how the inserted pauses related to markers of the musical phrase structure, such as changes in pitch height and tone duration.

General Discussion

The present series of experiments provided additional support for Krumhansl and Jusczyk's (1990) claims that infants are sensitive to certain aspects of musical phrase structure. In particular, the results of Experiments 1 and 2 undercut two potential alternative explanations for the Krumhansl and Jusczyk findings. Experiment 1 demonstrated that even when the naturally segmented and unnaturally segmented versions of the minuets began and ended in the same way, infants still preferred to listen to the versions that have pauses inserted at phrase boundaries (i.e., the naturally segmented versions) as opposed to in the middle of the phrases (i.e., the unnaturally segmented versions). Experiment 2 examined the issue of whether inserting pauses in the samples greatly altered the infants' perception of these pieces, rendering the minuets as more artificial than the unsegmented original versions. The results of this experiment suggested that this is not the case. The infants were no more inclined to listen to the unsegmented versions (original music) than they were to versions in which pauses were used to segment the music at the phrase boundaries (naturally segmented).

The remaining studies explored the issue of what cues might trigger the detection of musical phrase boundaries in infants. In the earlier study, Krumhansl and Jusczyk (1990) identified three musical variables that might play this sort of role in perception: pitch height, tone duration, and proportion of octave intervals. The results of Experiments 3, 4, and 5 helped to specify further the nature of the information to which infants are responding. In particular, Experiment 3 contrasted naturally segmented with reversed naturally segmented versions. The results indicated that the direction of pitch height changes and tone duration changes at phrase endings is critical. Infants listened significantly longer to the music played in the forward direction, in which pauses appeared just after the drops in pitch and the lengthening of durations of the tones in the melodic line. Experiment 4 demonstrated that mere discontinuities in the pitch height and tone duration variables, regardless of their direction, are not sufficient to produce the pattern of preferences; reversed naturally segmented versions were not preferred to reversed unnaturally segmented versions. Finally, the results of Experiment 5 showed that differences observed between reversed and forward versions of the minuets when pauses were inserted cannot be dismissed on the basis of claims that infants simply prefer the forward to the reversed versions; no preference was found for the original music over the reversed original music in which no pauses were inserted. Thus, the present study provides strong support for the view that musical phrase structure may be cued for infants by at least two of the musical variables that Krumhansl and Jusczyk (1990) suggested, namely, drops in pitch contour and increasing tone durations. As is discussed in more detail later, the third variable, the proportion of octave intervals, had less consistent effects.

As we have noted throughout, there are a number of apparent similarities between the cues that we are suggesting might underlie infants' segmentation of music into phrases and those hypothesized to underlie infants' segmentation of speech into clausal and phrasal units. For example, Hirsh-Pasek et al. (1987) used the same procedure as in the present study to demonstrate that infants as young as 6 months are sensitive to prosodic markers of clausal units in fluent speech. More recently, Jusczyk et al. (1992) showed that by 9 months of age American infants are sensitive to prosodic markers of phrasal units within clauses. Analyses of speech directed to children suggest that prosodic marking in the input is exaggerated relative to that directed to adults (e.g., Fernald & Simon, 1984; Fernald et al., 1989). In particular, a number of recent studies reported evidence for prosodic marking of phrasal boundaries in speech directed to children (e.g., Fisher, 1991; Jusczyk et al., 1992; Lederer & Kelly, 1992; Morgan, 1986). For instance, Lederer and Kelly reported consistent effects of syllable duration lengthening (the speech equivalent of increasing tone durations) at phrase boundaries in speech directed to children. Jusczyk et al. (1992) found some evidence of syllable lengthening but also very consistent effects of drops in fundamental frequency (the speech equivalent of drops in pitch contour) at the ends of phrasal units.

Given the apparent parallels in the marking of the speech and musical input and infants responses to these markers, it is tempting to conclude that the same processes are at work in both cases. However, closer scrutiny of the speech and musical analyses suggests that a more cautious approach is warranted. For example, consider the average size of the pitch drops in the present study (235 Hz on average, slightly more than an octave) in relation to those that Jusczyk et al. (1992) reported for their speech samples (25 Hz on average). The magnitude of the change in the speech samples is clearly much smaller than in the musical samples. This raises the question as to whether a change in the pitch contour of musical stimuli equivalent to the one for speech would still trigger the perception of musical phrase boundaries in infants. Vocal melodies, for example, typically exhibit smaller pitch variations than the piano music used here. Finally, although the present results give considerably more weight to the direction of change in the musical variables, as opposed to the discontinuity in these variables across successive phrases, it is interesting that Jusczyk et al. did not find evidence of a similar discontinuity across phrases in their speech samples. That is, although drops in pitch tended to occur at the ends of phrases in speech, the initial pitches of the next phrase did not immediately return to a higher level.

Our previous study (Krumhansl & Jusczyk, 1990) had suggested a third variable—the presence of octave intervals at phrase endings—that might also contribute to the infants' perception of phrase structure. In the present experiments, this variable had inconsistent effects. However, there were only two experiments in which the contrasted versions contained pauses, and they differed in terms of the number of octaves preceding the pauses. In Experiment 1 there was a significant positive correlation between the proportion of octave intervals and infants' orientation times. In Experiment 4, there was a significant correlation with orientation times but in a negative direction. In this experiment, however, the two versions did not differ in terms of the pitch height and tone duration variables, suggesting the octave variable may not operate in isolation from the two other cues to phrase structure. The following argument may explain why the octave variable may be a less potent cue to musical phrase structure. The variables of pitch and duration have a wide range of potential applications in auditory perception beyond musical stimuli. As noted, these cues appear to have parallels in speech perception and possibly in other domains of auditory perception. Consequently, it would not be surprising if infants were innately predisposed to attend to such cues to organize and group auditory information into its underlying event structure. By comparison, the octave interval variable involves a specific pitch relation that is likely to apply primarily in the domain of music.

The argument that the perceptual effects of musical intervals depend to some extent on experience with music is further supported by the cross-cultural differences found in tuning systems and scale structures (see, e.g., Krumhansl 1987, 1990). These differences suggest that intervallic structure is acquired, or at least shaped, by exposure to the culture's music. In support of this, research by Lynch, Eilers, Oller, and Urbano (1990) with 6-month-olds found that, in contrast to adults, American infants showed no special sensitivity to mistunings in a Western musical scale context. Instead, they were equally sensitive to a mistuned tone in a non-Western scale context as in a Western scale context. In a similar vein, Trainor and Trehub (1992) reported that, although adults find changes to a nondiatonic tone easier to detect than changes to a diatonic tone, infants find both types of change equally easy to detect. On the other hand, other results suggest that infants may be sensitive to certain musical intervals. Demany and Armand's (1984) results indicate that infants treat tones separated by octaves as equivalent. A study by Trehub, Thorpe, and Trainor (1990) found that infants of about 8½ months were better at detecting a one-semitone change in a melody based on a dominant-tonic progression compared with other melodies in either Western or non-Western tuning systems. Finally, Cohen, Thorpe, and Trehub (1987) demonstrated that infants have the ability to discriminate between melodies based on major, minor, and augmented chords. Moreover, they found changes from major to augmented chords were easier to detect than changes from augmented to major chords, suggesting the major chord has a special perceptual status. Clearly, more research is needed to determine which aspects of musical perception are universal and presumably innate and which aspects depend on learning about the structure of particular musical systems. The present studies suggest that basic grouping principles appear to be quite reliable and robust in infancy.

How might these basic grouping principles facilitate the acquisition of the particular intervallic structures of a musical culture? Krumhansl (1990) noted that in Western music the tones that are considered most structurally significant from a music-theoretical point of view tend to be sounded more frequently and for longer durations. So the tonic and dominant tones appear most frequently, followed by the other scale pitches, and nondiatonic tones are less frequent. Cross-cultural studies (Castellano, Bharucha, & Krumhansl, 1984; Kessler, Hansen, & Shepard, 1984) show that listeners unfamiliar with the style tend to use these cues for identifying the structurally important tones. Analogous results have been found for the unfamiliar Western style of atonality (Krumhansl, Sandell, & Sergeant, 1987). In a similar way, sensitivity on the part of infants to these distributional properties could aid them in acquiring knowledge of Western scale structure and harmony. Moreover, this process might be facilitated by the fact that phrases tend to end on the more structurally significant tones. Interestingly, an analysis of the musical samples used in Krumhansl and Jusczyk's (1990) study showed that 69% of the tones preceding the phrase endings were either the tonic or dominant of the key, and none were nondiatonic (nonscale) tones. Therefore, attention to the final tones would tend to select the most structurally significant tones in the Western tonal system. Thus, there may be an intimate connection between the early sensitivity to phrase structure demonstrated in this and the previous study and the acquisition of tonal structure.

In the same way, there are indications that the acquisition and development of language may depend on both innate abilities and knowledge of the structural properties of particular languages. For instance, studies of speech perception capacities of infants younger than 6 months suggest that the infants are capable of perceiving contrasts that could occur in any of the world's languages (e.g., see Aslin, Pisoni, & Jusczyk, 1983; Jusczyk, in press; Kuhl, 1987, for reviews). However, there are indications that sensitivity to certain nonnative speech contrasts begins to decline during the latter half of the first year (e.g., Werker & Lalonde, 1988; Werker & Tees, 1984; cf. Best, McRoberts, & Sithole, 1988). Moreover, during this same period, there are indications that infants are beginning to pick up information about the distributional properties of the elementary sounds used in their native language. Thus, Jusczyk, Friederici, Wessels, Svenkerud, and Jusczyk (1992) reported that 9-month-olds, but not 6-month-olds, are sensitive to constraints that govern which sequences of segments are permissible in words in their native language. Similarly, Jusczyk, Cutler, and Redanz (in press) found evidence that suggests that American 9-month-olds, but not 6-month-olds, have picked up information about the predominant stress patterns of frequently used words in English. Furthermore, it is almost certainly the case that some aspects of speech segmentation must also depend on the acquisition of knowledge of the structural characteristics of particular languages. To give just one example, prosodic marking of subclausal units such as phrases is apt to be much more prevalent and consistent in word-order languages (such as English) than in case languages (such as Polish).

In conclusion, the present study provides further evidence that 4½-month-olds are sensitive to certain aspects of musical phrase structure. In addition, there are indications that drops in pitch contour and increases in tone durations are the musical variables that may serve to mark musical phrase endings for infants. Nevertheless, considerable work remains in order to arrive at a full understanding of the way in which the perception of musical phrase structure develops. To this point, our studies have made no effort to isolate the pitch height cue from the tone duration cue. However, if we are to understand the relative contributions that these cues make to segmenting the musical passages, then we must begin to isolate these cues and observe the role that each plays in infants' perception of musical phrase endings. For example, if either pitch height or tone duration were held constant, would infants still display the same pattern of preferences for the musical stimuli? Another important issue to address concerns the extent to which the kinds of preferences that we have observed with these simple Mozart minuets will also hold for more music in other styles. Finally, there is the issue that concerns the generality of the cues to which infants are responding. Are there general properties of acoustic signals that serve to mark salient units in speech, music, and other auditory domains, or do different mechanisms subserve segmentation in each of these domains?

References

Antell, S. E., & Keating, D. (1983). Perception of numerical invariance by neonate. *Child Development, 54,* 695–701.

Aslin, R. N., Pisoni, D. B., & Jusczyk, P. W. (1983). Auditory development and speech perception in infancy. In M. M. Haith & J. J. Campos (Eds.), *Infancy and the biology of development* (pp. 573–687). New York: Wiley.

Baillargeon, R., Graber, M., DeVos, J., & Black, J. (1990). Why do young infants fail to search for hidden objects. *Cognition, 36,* 255–284.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of the perceptual re-organization for speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Jour-*

nal of Experimental Psychology: Human Perception and Performance, 14, 245-360.

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

Chang, H. W., & Trehub, S. E. (1977). Auditory processing of relational information by young infants. *Journal of Experimental Child Psychology, 24,* 324–331.

Clarke, E. F., & Krumhansl, C. L. (1990). Perceiving musical time. *Music Perception, 7,* 213–252.

Cohen, A. J., Thorpe, L. A., & Trehub, S. E. (1987). Infants' perception of musical relations in short transposed tone sequences. *Canadian Journal of Psychology, 41,* 33–47.

Cooper, W., & Sorensen, J. (1977). Fundamental frequency contours at syntactic boundaries. *Journal of the Acoustical Society of America, 62,* 683–692.

Crystal, T. H., & House, A. S. (1988). Segmental durations in connected speech signals: Current results. *Journal of the Acoustical Society of America, 88,* 101–112.

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

Deliege, I. (1989). Recognition of musical forms through listening. *Contemporary Music Review, 14,* 214–230.

Demany, L. (1982). Auditory stream segregation in infancy. *Infant Behavior and Development, 5,* 261–276.

Demany, L., & Armand, F. (1984). The perceptual reality of tone chroma in early infancy. *Journal of the Acoustical Society of America, 76,* 57–66.

Deutsch, D., & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. *Psychological Review, 88,* 181–195.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 8,* 181–195.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27,* 209–221.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20,* 104–113.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16,* 477–501.

Fisher, C. L. (1991, April). *Prosodic cues to phrase structure in infant directed speech.* Paper presented at the Stanford Child Language Research Forum, Stanford, CA.

Haith, M. M., & Campos, J. J. (Eds.). (1983). *Infancy and developmental psychobiology* (4th ed.). New York: Wiley.

Hirsh-Pasek, K., Kemler Nelson, D. G., Juscyzk, P. W., Wright Cassidy, K., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition, 26,* 269–286.

Hockett, C. A. (1954). Two models of grammatical description. *Word, 10,* 210–231.

James, W. (1890). *The principles of psychology.* New York: Holt.

Jusczyk, P. W. (1989, April). *Perception of cues to clausal units in native and non-native languages.* Paper presented at the biennial meeting of the Society for Research in Child Development, Kansas City, MO.

Jusczyk, P. W. (in press). Developmental speech perception. In N. J. Lass (Ed.), *Principles of experimental phonetics.* Philadelphia: Decker.

Jusczyk, P. W., Cutler, A., & Redanz, N. (in press). Infants' preference for predominant stress patterns of English words. *Child Development.*

Jusczyk, P. W., Friederici, A. D., Wessels, J., Svenkerud, V. Y., & Jusczyk, A. M. (1992). *Infants' sensitivity to the sound patterns of native language words.* Manuscript submitted for publication.

Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology, 24,* 252–293.

Kellman, P. J., Gleitman, H., & Spelke, E. S. (1987). Object and observer motion in the perception of objects by infants. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 586–593.

Kessler, E. J., Hansen, C., & Shepard, R. N. (1984). Tonal schemata in the perception of music in Bali and the West. *Music Perception, 2,* 131–165.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in connected discourse. *Journal of Phonetics, 3,* 129–140.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics, 7,* 279–312.

Krumhansl, C. L. (1987). General properties of musical pitch systems: Some psychological considerations. In J. Sundberg (Ed.), *Harmony and tonality.* Stockholm: Royal Swedish Academy of Music.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch.* New York: Oxford University Press.

Krumhansl, C. L., & Jusczyk, P. W. (1990). Infants' perception of phrase structure in music. *Psychological Science, 1,* 70–73.

Krumhansl, C. L., Sandell, G. J., & Sergeant, D. C. (1987). The perception of tone hierarchies and mirror forms in twelve-tone serial music. *Music Perception, 5,* 31–78.

Kuhl, P. K. (1987). Perception of speech and sound in early infancy. In P. Salapatek & L. Cohen (Eds.), *Handbook of infant perception* (pp. 275–381). New York: Academic.

Lederer, A., & Kelly, M. H. (1991). Prosodic correlates to the adjunct/complement distinction in motherese. In *Papers & Reports on Child Language Development, 30,* Stanford, CA.

Lederer, A., & Kelly, M. (1992). *Prosodic information for syntactic structure in parental speech.* Manuscript submitted for publication.

Lehiste, I., Olive, J. P., & Streeter, L. (1976). The role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America, 60,* 1199–1202.

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

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition, 25,* 265–288.

Lynch, M. P., Eilers, R. E., Oller, D. K., & Urbano, R. C. (1990). Innateness, experience, and music perception. *Psychological Science, 1,* 70–73.

Mehler, J., & Fox, R. (Eds.). (1985). *Neonate cognition: Beyond the blooming, buzzing confusion.* Hillsdale, NJ: Erlbaum.

Morgan, J. L. (1986). *From simple input to complex grammar.* Cambridge, MA: MIT Press.

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

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

Piaget, J. (1954). *The construction of reality in the child.* New York: Basic Books.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustic Society of America, 90,* 2956–2970.

Salapatek, P., & Cohen, L. (Eds.). (1987). *Handbook of infant perception.* New York: Academic Press.

Scott, D. R. (1982). Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America, 71,* 996–1007.

Scott, D. R., & Cutler, A. (1984). Segmental phonology and the perception of syntactic structure. *Journal of Verbal Learning and Verbal Behavior, 23,* 450–466.

Simon, H. A., & Sumner, R. K. (1968). Pattern in music. In B. Kleinmuntz (Eds.), *Formal representation of human judgement.* New York: Wiley.

Spelke, E. S. (1985). Perception of unity, persistence, and identity: Thoughts on infants' conceptions of objects. In J. Mehler & R. Fox (Eds.), *Neonate cognition: Beyond the blooming buzzing confusion* (pp. 89–113). Hillsdale, NJ: Erlbaum.

Starkey, P., Spelke, E. S., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition, 36,* 97–127.

Streeter, L. A. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America, 64,* 1582–1592.

Thorpe, L. A., & Trehub, S. E. (1989). Duration illusion and auditory grouping in infancy. *Developmental Psychology, 24,* 484–491.

Thorpe, L. A., Trehub, S. E., Morrongiello, B. A., & Bull, D. (1989). Perceptual grouping by infants and preschool children. *Developmental Psychology, 24,* 484–491.

Trainor, L. J., & Trehub, S. E. (1992). A comparison of infants' and adults' sensitivity to Western musical structure. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 394–402.

Trehub, S. E. (1987). Infants' perception of musical patterns. *Perception & Psychophysics, 41,* 635–641.

Trehub, S. E., Thorpe, L. A., & Trainor, L. J. (1990). Infants' perception of good and bad melodies. *Psychomusicology, 9,* 5–19.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology, 24,* 672–683.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49–63.

Received April 29, 1992
Revision received September 15, 1992
Accepted September 16, 1992 ■