

# Learning and Discrimination of Audiovisual Events in Human Infants: The Hierarchical Relation Between Intersensory Temporal Synchrony and Rhythmic Pattern Cues

David J. Lewkowicz

New York State Institute for Basic Research in Developmental Disabilities

This study examined 4- to 10-month-old infants' perception of audio-visual (A-V) temporal synchrony cues in the presence or absence of rhythmic pattern cues. Experiment 1 established that infants of all ages could successfully discriminate between two different audiovisual rhythmic events. Experiment 2 showed that only 10-month-old infants detected a desynchronization of the auditory and visual components of a rhythmical event. Experiment 3 showed that 4- to 8-month-old infants could detect A-V desynchronization but only when the audiovisual event was nonrhythmic. These results show that initially in development infants attend to the overall temporal structure of rhythmic audiovisual events but that later in development they become capable of perceiving the embedded intersensory temporal synchrony relations as well.

A ticking metronome, a tap dancer, and a talking person all illustrate the fact that many events in our everyday world are specified concurrently in multiple sensory modalities. In addition, the multimodal sensory information that specifies them is distributed over time. The specific way the information is distributed (i.e., its temporal structure) determines the perceptual and cognitive meaning of temporally defined events (Baldwin & Baird, 2001; Zacks & Tversky, 2001). The two best examples of the fundamental importance of temporal structure for perception and cognition are, of course, music and language. In each case, the particular temporal organization of a series of elements, be they notes or phonemes, can give rise to very different meanings (Bregman, 1990; Fraisse, 1982a, b; Krumhansl, 2000; Lashley, 1951; Martin, 1972; Pomerantz & Lockhead, 1991).

In general, empirical evidence indicates that infants are sensitive to temporal structure in both the auditory and visual modalities (Lewkowicz, 1989, 2000a). For example, it has been reported that infants can perceive the temporal organization of a sequence of identical (Demany, McKenzie, & Vurpillot, 1977) or distinct (Chang & Trehub, 1977) sounds, that they can detect changes in the duration of the silent intervals that separate sounds (Thorpe &

Trehub, 1989; Thorpe, Trehub, Morrongiello, & Bull, 1988), and that they can discriminate between different visual rhythms (Mendelson, 1986). Evidence also suggests that whereas sensitivity to some forms of temporally distributed sensory input remains unchanged, sensitivity to other forms of this type of input improves with development. Thus, sensitivity to audio-visual (A-V) synchrony relations and the ability to discriminate audiovisual rate variations emerge early and remain unchanged throughout infancy (Lewkowicz, 1992b, 1996). In contrast, thresholds for the detection of auditory gaps decrease with age (Trehub, Schneider, & Henderson, 1995; Werner, Marean, Halpin, Spetner, & Gillenwater, 1992), and the ability to discriminate more complex acoustic rhythms improves with age (Morrongiello, 1984).

Usually, when we think of temporal structure, we think of rhythm. Fraisse (1982b) defined *rhythm* as an ordered succession of elements that can be temporally distributed in either a regular or an irregular fashion. The sound of a ticking metronome exemplifies a regularly distributed sequence and thus constitutes what Fraisse referred to as an *isochronous* sequence. A Mozart minuet, on the other hand, exemplifies what we usually think of as a rhythmic pattern. Its primary characteristic is that its constituent elements are separated by unequal intervals of time. This unequal temporal distribution leads to perceptual chunking of various groups of sounds into distinct, Gestalt-like groupings that have definite beginnings and ends. Martin (1972) agreed with Fraisse's distinction between isochronous and rhythmic sequences and made the additional critical point that only patterned rhythmical sequences can be characterized in terms of relative timing differences. By *relative timing*, Martin (1972) meant that "the locus of each (sound) element along the time dimension is determined relative to the locus of all other elements in the sequence, adjacent and nonadjacent" (p. 488).

Applying the relative timing criterion to the study of rhythm perception means that the different rhythmic patterns used in a given study of discrimination must differ in terms of the relative arrangement of the intervals separating each element of a pattern. For example, a 2-2 rhythmic pattern of hammer taps would consist of four taps separated, in turn, by short, long, and short intervals,

---

David J. Lewkowicz, New York State Institute for Basic Research in Developmental Disabilities.

This work was supported in part by funds from the New York State Office of Mental Retardation and Developmental Disabilities and in part by National Institute of Child Health and Human Development Grants R03 HD36731 and R01 HD35849.

I thank Marcia Dabbene and Linnea Dickson for their assistance. I also express my gratitude to Lorraine Bahrick, Robert Lickliter, and Stuart Marcovitch for helpful discussions regarding this work and for useful comments and suggestions on an earlier version of the manuscript.

Correspondence concerning this article should be addressed to David J. Lewkowicz, who is now at the Department of Psychology, Charles E. Schmidt College of Science, Florida Atlantic University, E&S Building, 2912 College Avenue, Davie, Florida 33314. E-mail: lewkowicz@fau.edu

whereas a 3–1 pattern would simply involve a rearrangement of the intervals: short, short, and long. Although the relative timing criterion ensures that discrimination is based on the overall temporal organization of the sequence, one other requirement must be met to ensure against possible confounds. The elements comprising different rhythmic patterns should be identical. This constant feature criterion means that if the patterns are auditory, then their elements should not differ in pitch, timbre, or some other acoustic feature, and if the patterns are visual, then their elements should not differ in color, shape, or some other visual feature.

The problem with many of the extant studies on infants' response to rhythmic structure is that they do not meet the relative timing and/or the constant feature criteria. For example, the study of acoustic pattern discrimination by Demany et al. (1977), which is often cited as evidence of rhythm discrimination in 2-month-old infants, illustrates a violation of the relative timing criterion. Specifically, the rhythmic pattern differences in one of the experiments were confounded with absolute differences in the interelement intervals, making it possible that discrimination was based on the absolute duration of the interelement intervals rather than on the overall pattern difference. In another experiment, Demany et al. (1977) manipulated the relative order of interelement intervals, but there the different rhythmic sequences were composed of only three elements. As a result, successful discrimination might have been based on the first interelement interval rather than the overall pattern. A study by Mendelson (1986) of infant discrimination of visual rhythmic patterns suffers from similar confounds. Although the theoretical justification for the relative timing and the constant feature criteria is self-evident, Morrongiello (1984) actually provided empirical evidence that relative timing does make a difference. She compared discrimination of rhythmic patterns (composed of a series of identical sounds) based on either absolute or relative differences in their interelement intervals. She found that 6- and 12-month-old infants successfully discriminated between different rhythmic patterns when their interelement intervals differed in absolute duration but that only 12-month-old infants discriminated between rhythmic patterns that differed in the relative ordering of their intervals.

Some of the clearest evidence that infants can perceive the Gestalt-like quality of temporal structure comes from studies of infants' discrimination of bimodal, audiovisual rhythmic patterns (for definitions of terms related to multimodal perception, please see the Appendix). These studies have shown that infants as young as 5 months of age can discriminate between relatively complex bimodal rhythms (Bahrck & Lickliter, 2000; Pickens & Bahrck, 1995, 1997). Of particular interest is Bahrck and Lickliter's (2000) finding that infants can successfully learn and discriminate between different rhythmic patterns but only if the patterns are specified concurrently by temporally synchronous audible and visible attributes during the learning phase. The finding that A-V temporal synchrony is essential for successful learning and discrimination of rhythm is consistent with many other findings. For example, infants exhibit more effective learning when the auditory and visual attributes of an audiovisual event are synchronous (Bahrck, 1988), and they prefer to look at a talking face whose voice is synchronized with it rather than at a talking face whose voice is not (Dodd, 1979). Likewise, infants prefer to look at a bouncing object that is synchronized with a sound rather than at a bouncing object that is not synchronized with a sound (Lewko-

wicz, 1992a; Spelke, 1979). Finally, infants perform cross-modal matching of the duration of auditory and visual information but only if the information in the two modalities is temporally synchronous (Lewkowicz, 1986).

The importance of temporal synchrony for the perception of intersensory relations has clear face validity. The everyday perceptual world is full of multimodal objects and events. The auditory, visual, tactile, and olfactory features of such objects and events are usually available at the same time and usually also at the same place. For example, regardless of whether it is a person talking, a ball bouncing, or a steak sizzling on a grill, one can simultaneously hear, see, and, in the case of the steak, smell these objects or events. Not surprisingly, empirical evidence indicates that adults are sensitive to these kinds of intersensory temporal synchrony relations (Dixon & Spitz, 1980; Massaro, 1998; McGrath & Summerfield, 1985; Radeau, 1994). Moreover, findings from positron emission tomography studies in human adults and anatomical and electrophysiological findings from animal studies indicate that the detection of intersensory temporal synchrony is mediated by a low-level, subcortical network consisting of the thalamus, superior colliculus, and the right insula region of cortex (Bushara, Grafman, & Hallett, 2001). This network makes it possible to detect the temporal correspondence of auditory and visual inputs at an early stage of cortical processing. Findings such as these suggest that the neural mechanisms underlying the detection of A-V synchrony relations are basic and most likely phylogenetically old. In addition, as pointed out earlier, mechanisms for the detection of intersensory temporal synchrony are functional early in development.

Given that infants are sensitive to intersensory temporal synchrony from an early age and that this perceptual cue usually accompanies other intersensory temporal cues, it is not surprising that synchrony plays a role in the perception of multimodal rhythm. Indeed, Bahrck and Lickliter's (2000) findings support this conclusion by showing that infants can discriminate bimodal rhythms only if their auditory and visual attributes are synchronous. What is interesting, however, is that intersensory temporal synchrony cues might be hierarchically related to rhythmic pattern cues in a situation in which both are available concurrently. That is, it is possible that synchrony cues might play a supportive role when they accompany rhythmic pattern cues by making the rhythmic pattern perceptually more salient than it would be if the audible and visible attributes were not synchronous.

One of the principal reasons why intersensory temporal synchrony cues might play a subordinate role to rhythmic pattern cues is that the latter impose structure on sequential sensory input and thus are much more perceptually compelling. Indeed, the human propensity to perceive temporal structure is so strong that our perceptual systems perform what Fraise (1982b) called "subjective rhythmization" *even* on sequential input that is not temporally structured (i.e., when it is isochronous). In other words, our perceptual systems appear to be predisposed to perceiving the temporal world around us in terms of temporal chunks. If it is assumed that the propensity for temporal chunking is present early in development, then it is especially likely that young, perceptually inexperienced infants might not be capable of detecting synchrony cues in their own right when they co-occur with rhythmic pattern cues. This may be because rhythmic pattern cues may simply overwhelm the young infants' perceptual systems to such a point

that infants do not detect the intersensory temporal synchrony aspects of a bimodal stimulus. As they become more perceptually skilled, however, infants are likely to become capable of detecting synchrony as well as rhythmic pattern cues. The purpose of this study was to test this developmental hypothesis.

Because a test of the hypothesis first required information on infants' discrimination of rhythm across a broad age range, in Experiment 1 I tested 4-, 6-, 8-, and 10-month-old infants to provide baseline data on discrimination of bimodally synchronous rhythmic patterns. Thus, infants were habituated to one of two different patterns that met the relative timing and constant feature criteria and that consisted either of a person uttering a syllable in a rhythmic manner or of a toy hammer tapping rhythmically. Then, in Experiments 2 and 3, I tested the hypothesis directly by investigating infants' detection of A-V desynchronization in both a rhythmic (Experiment 2) and a nonrhythmic (Experiment 3) context.

## Experiment 1

### Method

**Participants.** There were two groups of 4-, 6-, 8-, and 10-month-old infants: a syllable group, which consisted of a total of 137 infants at these four ages, and a hammer group, which consisted of a total of 175 infants at these four ages. All the infants tested in this experiment, as well as those in the two subsequent ones, were full-term at the time of birth, had 1- and 5-min Apgar scores of 7 or higher, and were healthy at the time of testing.

The 4-month-old syllable group (mean age = 19.2 weeks,  $SD = 0.8$  week) consisted of 19 boys and 19 girls, the 6-month-old group (mean age = 27.7 weeks,  $SD = 0.8$  week) consisted of 25 boys and 9 girls, the 8-month-old group (mean age = 36.3 weeks,  $SD = 0.6$  week) consisted of 19 boys and 14 girls, and the 10-month-old group (mean age = 44.8 weeks,  $SD = 1.0$  week) consisted of 16 boys and 16 girls. An additional 14 syllable-group infants were tested but did not contribute usable data because of fussing, sleepiness, or experimenter error.

The 4-month-old hammer group (mean age = 19.3 weeks,  $SD = 0.9$  week) consisted of 22 boys and 21 girls, the 6-month-old group (mean age = 28 weeks,  $SD = 0.9$  week) consisted of 30 boys and 15 girls, the 8-month-old group (mean age = 36.4 weeks,  $SD = 0.8$  week) consisted of 24 boys and 23 girls, and the 10-month-old group (mean age = 45 weeks,  $SD = 1.1$  weeks) consisted of 24 boys and 16 girls. An additional 12 hammer-group infants were tested but did not contribute usable data because 3 of them fussed, 4 were inattentive, and 5 could not be tested because of equipment failure.

**Apparatus and stimuli.** Depending on their age and/or willingness to sit alone, infants were seated in either an infant seat, a high chair, or a parent's lap in front of a 13-in. (33-cm) color video monitor (Panasonic CT-1331Y) located at a distance of approximately 50 cm. Infants who were fussy when first placed in the infant seat or high chair were placed on the parent's lap (22 infants in the syllable group and 36 infants in the hammer group were tested in this way). The person holding the infant was unaware of the specific purpose of the experiment and was asked to sit as still as possible and refrain from interacting with the baby. A color camcorder (Panasonic AG-190) that was located on top of the stimulus display monitor was used to videotape and view the infant's behavior. A curtain that extended from the monitor past the point where the infant was seated was placed on each side to occlude the infant's peripheral view. The experimenter was seated on the other side of the curtain and watched the infant on a monitor while listening to continuous broadband noise through a set of headphones. In this way, the experimenter could neither see nor hear the stimuli presented to the infant.

Six different events were constructed and turned into multimedia movies for the syllable group. One of these movies served as an "attention getter"

and showed a silent green disk that alternately expanded and shrank (measuring 8 cm in diameter at its largest size and 2 cm at its smallest size). A second movie consisted of a segment of a Winnie-the-Pooh cartoon and was used to test for fatigue effects at the beginning and end of the testing session. The four remaining movies were the habituation/test movies. They were recordings of one of two female actors uttering the syllable /ba/ in a rhythmic fashion. Construction of the rhythmic syllable patterns consisted of first filming each actor while she repeatedly uttered the syllable /ba/. During the filming she held her head still, recited the syllable with a minimum amount of intonation, and did not smile. The top and bottom of each actor's head corresponded to the top and bottom edges of the stimulus display monitor. This initial video recording was then reviewed, and a single instance of the syllable was selected to represent each actor, with the proviso that a short segment of still face preceding and following the syllable's utterance be visible. The duration of one actor's syllable (measured from onset to offset of phonation) was 300 ms, whereas the duration of the other actor's syllable was 330 ms.

The two syllables selected to represent each actor were used to construct two different rhythmic patterns with Adobe Corporation's Premiere video-editing program. The two rhythmic patterns consisted of the repetition of the syllable four times according to a 2-2 or a 3-1 temporal pattern. Specifically, a single rhythmic cycle of each pattern consisted of four repetitions of the identical syllable followed by a silent interval that was longer than any of the intersyllable intervals. The overall duration of the rhythmic cycle (which included a 1,900-ms silent interval) was 3,800 ms. To prevent infants from basing their discriminative responses on some local temporal feature of the temporal sequence and to force them to process the Gestalt property of the temporal sequences, the 2-2 and 3-1 rhythmic patterns differed only in terms of the relative position of the intervals that separated each of the syllables. In other words, the initial intersyllable interval was identical across the two rhythmic patterns, and the difference between them was internal to the patterns. Represented in terms of X's (syllables) and O's (rests), the 2-2 rhythmic pattern was an X X O X X grouping of syllables and rests, whereas the 3-1 rhythm was an X X X O X grouping of syllables and rests. There were two types of intersyllable intervals: the short one, separating the X X part of the pattern, and the long one (the rest). The short one ranged between 200 and 250 ms, whereas the long one ranged between 500 and 550 ms.

In producing the rhythmic patterns, the aim was to make the transitions from one instance of the syllable to the next as continuous and seamless as possible. To do so, I used the cross-dissolve transition effect in Premiere to concatenate the four identical copies of the selected syllable. The cross-dissolve effect began during the still face portion at the end of one syllable and continued into the beginning of the still face portion of the next instantiation of the syllable. In this way, the effect made the transition essentially imperceptible and gave an observer the impression of a person repeatedly uttering a syllable in a natural fashion. Once a single 2-2 or 3-1 rhythmic pattern was constructed, Adobe's Premiere was then used to assemble 13 instances of the rhythmic groupings into a 50-s segment that was, in turn, rendered into a QuickTime movie. While the movie was being rendered, a low-pass filter set to 4019 Hz was used to remove background noise.

The identical video-editing procedures were used for making the rhythmic patterns for the hammer group. First, a video recording of a blue plastic toy hammer tapping repeatedly against a visible surface was made. The hammer and its handle were visible but the hand holding the hammer was not. Then a single instance of this action was chosen for construction of the two different rhythmic patterns. The single tap (lasting 100 ms) consisted of the hammer suspended in the horizontal position over the surface, its downward movement, a tap, and an upward movement back to the horizontal position. By following the same procedures used to make the syllable rhythmic patterns, I used the single copy of the hammer tap to construct the two hammer rhythmic patterns and then the movies of each pattern. The short intertap interval, calculated from tap to tap, was 561 ms,

the long intertap interval was 792 ms, and the silent interval between the four taps was 1,881 ms. The same attention-getter and cartoon movies that were used for the syllable group were used for the hammer group.

Testing was done with either a personal computer (PC) or a Macintosh IIfx computer. The PC presented the movies online, whereas the Macintosh computer presented the movies by controlling a Sony LDP-1550 laser-disk player that played the movies, which had previously been pressed onto a laser disk. In addition to controlling all stimulus presentations, each computer measured and recorded the duration of infants' visual fixations during stimulus presentation.

Average sound pressure level (SPL) of the audio component of the movies was measured with a Brüel and Kjær (Norcross, GA) 2203 sound-level meter (A scale, re .0002 dynes/cm<sup>2</sup>) with the microphone placed at the same place as the infant's head. The audio portion of the stimulus events presented on the PC had the following average SPLs: 65–70 dB for the cartoon, 70 dB for the syllable spoken by one actor, 73 dB for the syllable spoken by the other actor, and 69 dB for the hammer tap. The average SPLs of the stimulus events presented on the Macintosh computer were 68–71 dB for the cartoon, 71 dB for the syllable spoken by one actor, 70 dB for the syllable spoken by the other actor, and 62 dB for the hammer tap.<sup>1</sup>

*Procedure and design.* An infant-controlled habituation/test procedure was used. This meant that the infant's looking at the stimulus monitor during a given experimental trial controlled stimulus presentation and thus the duration of each trial. Specifically, each time the infant looked at the stimulus monitor, the experimenter pressed either a key on the computer keyboard or a mouse button attached to the computer. This initiated the presentation of the appropriate stimulus, which continued for as long as the infant continued to look at the monitor. Whenever the infant looked away from the monitor, the experimenter stopped pressing the key or mouse button. If the infant looked back at the stimulus within 1 s, the presentation of the stimulus (and thus the trial) continued. The trial ended only if one of two conditions was met: Either the infant did not look back within 1 s, or 34 s elapsed. Once the infant either met the look-away criterion or reached the maximum trial duration, the attention-getter movie reappeared on the monitor. As soon as the infant looked at the attention-getter movie, it disappeared, and the experimenter initiated the next trial by presenting the appropriate stimulus event. Trials continued in this fashion until the infant reached a habituation criterion of a 50% or greater decline in looking duration. The criterion was calculated by comparing the total duration of looking during the last three habituation trials to the total duration of looking during the first three habituation trials. A sliding window was used for the computation of the habituation criterion in that the calculation of the duration of looking for the last three trials of the habituation phase began with the second trial and continued to slide down a trial at a time until the habituation criterion was reached. Once an infant reached the habituation criterion, the test phase began without interruption and continued until all remaining trials were presented.

The experiment commenced with the presentation of a single pretest trial in which the cartoon segment was presented. The purpose of this trial was to measure the initial level of attention. Once this trial ended, the habituation phase commenced, and infants in each group saw and heard the appropriate stimulus. At each age, infants in the syllable group were assigned to one of four habituation subgroups generated by crossing each of the two actors with each of the two rhythms. In the hammer group, infants at each age were assigned to one of two habituation rhythm subgroups.

Once the habituation phase ended, each infant was given two test trials. One of these was a familiar test trial in which the same rhythmic pattern that was presented during the habituation phase was presented again, whereas the other test trial was a novel one in which the novel rhythmic pattern was presented. The order of these two test trials was counterbalanced across infants in each age group. Looking during the familiar test trial served as a baseline against which the duration of looking to the novel

rhythm was assessed to determine if significant response recovery, and thus discrimination, occurred. Finally, a posttest trial was given in which the same cartoon that was presented during the pretest trial was presented again to measure the level of attention at the end of the experimental session.

## Results and Discussion

A preliminary repeated measures analysis of variance (ANOVA) was performed to determine whether the specific habituation stimulus, the infants' age, and/or fatigue affected responsiveness. The between-subjects factors for this analysis were age (4) and habituation stimulus (6), and the within-subjects factor was trial type (pretest, familiar, novel, and posttest). The dependent measure was the duration of looking. Results indicated that the specific stimulus/rhythm presented during the habituation phase did not affect responsiveness in that neither the main effect of habituation stimulus nor the interactions between this factor and the other two factors were statistically significant. The planned comparison tests between the pre- and posttest trials and between the familiar and posttest trials indicated that infants did not become fatigued during the experimental session. Specifically, there was no difference in the duration of looking in the pre- and posttest trials, and infants exhibited significant response recovery to the cartoon at the end of the testing session,  $F(1, 288) = 431.6$ ,  $p < .01$ .

Given that the specific type of habituation stimulus did not affect responsiveness, the data were collapsed over the specific habituation rhythm in each group and over the specific actor in the syllable group, yielding a grouping factor with two levels (hammer and syllable). A new repeated measures ANOVA was then computed on only the data from the familiar and novel test trials, and it included age (4) and group (2) as between-subjects factors and trial type (2) as a within-subjects factor. Results of this analysis yielded a significant age effect,  $F(3, 304) = 3.84$ ,  $p < .05$ , which was due to an age-based decline in the overall amount of time that infants spent looking regardless of what stimulus was presented. Age did not, however, interact with the other two factors. Of primary interest was the finding that the trials factor was significant,  $F(1, 304) = 60.3$ ,  $p < .01$ . This effect, plus the results from the posttest trial, can be seen in Figure 1. As can be seen, infants were highly engaged in the experimental task and discriminated the rhythms. In addition, the lack of any other significant interactions indicates that infants perceived the difference between the familiar and novel rhythms regardless of which specific rhythmic pattern they were habituated to and regardless of whether this pattern was produced by a talking person or by a rhythmically moving object.<sup>2</sup>

To explore the results further, I analyzed detection of rhythmic pattern differences separately at each age. Planned comparison

<sup>1</sup> The SPL of the hammer event played on the Macintosh computer was 7 dB lower than the SPL of the hammer event played on the PC. To make sure that this difference did not affect outcome, I reran the initial analysis outlined in the Results section with test computer included as a separate factor. No significant effects of test computer were found.

<sup>2</sup> To determine whether testing on the parent's lap affected outcome, I eliminated the data from all infants who were seated on a parent's lap and reanalyzed the data. This analysis yielded results identical to those obtained in the overall analyses.

tests comparing the duration of looking in the familiar and novel test trials were conducted on the combined data for the two groups but separately at each age. As can be seen in Figure 2, infants exhibited significant response recovery at each of the four ages,  $F_s(1, 304) = 31.7$  for the 4-month-olds, 9.27 for the 6-month-olds, 12.53 for the 8-month-olds, and 11.4 for the 10-month-olds (all  $ps < .01$ ).

The results from this experiment showed that at each of the ages tested, infants discriminated between 2-2 and 3-1 rhythmic patterns. The ability to make this discrimination was affected neither by whether the rhythmic pattern consisted of a sequence of syllables or a sequence of hammer taps nor, in the case of the syllables, by the specific person who uttered the syllable. Finally, consistent with findings from adults' perception of rhythmic structure (Frasse, 1982b), infants' perception of rhythm was not affected by the specific duration of the elements that made up the rhythmic patterns. That is, infants discriminated between the rhythmic patterns equally well regardless of whether the elements making up a particular rhythmic pattern had a duration of 100 ms (the hammer tap) or a duration of 300 or 330 ms. In other words, infants perceived the bimodally specified rhythmic patterns primarily on the basis of their overall temporal structure rather than on the basis of the specific temporal parameters of the elements.

Successful detection of the rhythmic pattern differences required that infants perceive changes in the relative position of identical interelement intervals. Moreover, discrimination could not have been based on the first interelement interval because this interval was the same in the two patterns. As a result, the current results show that infants perceived the Gestalt character of the sequences. As noted earlier, studies of infants' discrimination of acoustically specified rhythms involving relative interelement interval changes have shown that only 12-month-old infants could make such discriminations (Morrongiello, 1984). The fact that in the current experiment infants as young as 4 months of age could make a discrimination based on relative interelement changes suggests that bimodally specified rhythmic temporal structure is perceptually more salient and that this greater salience makes it possible for infants to detect rhythmic pattern differences earlier in development.

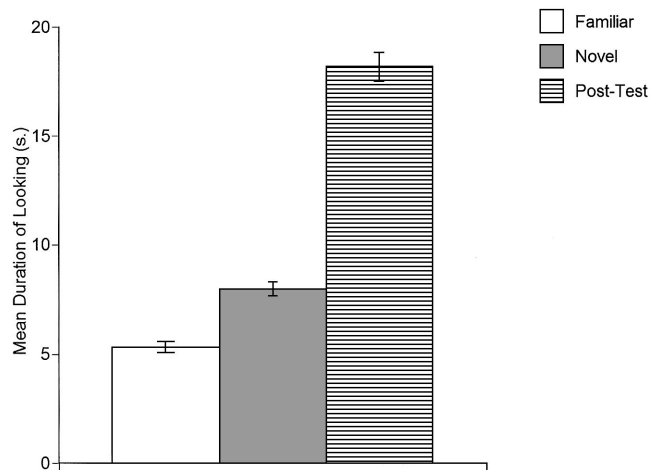


Figure 1. Mean duration of looking in the familiar, novel, and posttest trials in Experiment 1. Vertical lines depict standard errors of the means.

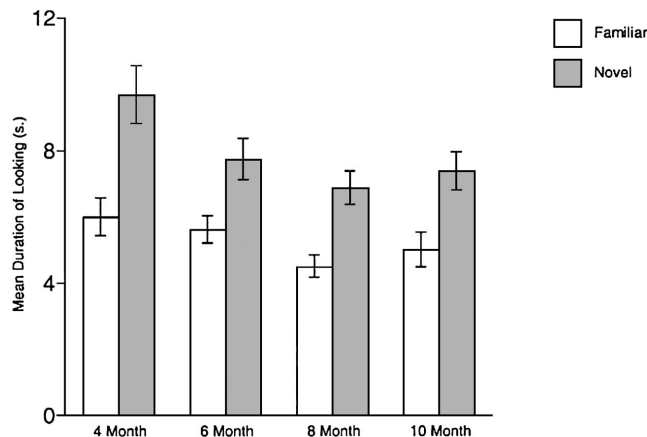


Figure 2. Mean duration of looking in the familiar and novel test trials in Experiment 1 as a function of age. Vertical lines depict standard errors of the means.

### Experiment 2

As noted in the introduction, the principal question of interest in the current study was how intersensory temporal synchrony cues interact with temporal pattern cues in infants' learning and discrimination of bimodal events. Having established in Experiment 1 that infants can learn and discriminate bimodally specified rhythms, in Experiment 2, I tested infants' ability to detect a desynchronization of the audible and visible components of a rhythmic pattern. Because the specific type of stimulus presented in Experiment 1 did not affect outcome, in the current experiment, infants were tested only with the syllable rhythmic patterns.

### Method

**Participants.** A total of 128 infants participated in this experiment. The sample consisted of separate groups of 4-, 6-, 8-, and 10-month-old infants, with 32 infants tested at each age. The 4-month-old group (mean age = 19.8 weeks,  $SD = 1.8$  weeks) consisted of 19 boys and 13 girls, the 6-month-old group (mean age = 28 weeks,  $SD = 0.9$  week) consisted of 19 boys and 13 girls, the 8-month-old group (mean age = 36.4 weeks,  $SD = 0.9$  week) consisted of 16 boys and 16 girls, and the 10-month-old group (mean age = 45.1 weeks,  $SD = 0.8$  week) consisted of 17 boys and 15 girls. Thirty-one infants refused to sit in the infant seat or in the high chair and thus were tested on the parent's lap. An additional 14 infants were tested but did not contribute usable data because 11 of these infants fussed, 1 was sleepy, and 2 were inattentive.

**Apparatus and stimuli.** The apparatus and stimuli were identical to those used in Experiment 1 except that only the syllable rhythmic patterns were presented. Infants were habituated to one of the four possible Actor  $\times$  Rhythm patterns. Following habituation to the synchronous syllable, infants were given three test trials. Two of these were the familiar test trial, in which the familiar rhythm was presented again, and a novel test trial, in which the familiar rhythm was presented again but this time with its audible and visible components desynchronized by 660 ms. The third test trial was the posttest trial, in which the cartoon was presented. The desynchronized novel stimulus was constructed in Adobe's Premiere by making the onset of the audible attribute of the syllable precede the onset of the visible attribute by 660 ms. This desynchrony value was chosen on the basis of prior studies showing that infants can detect this level of A-V desynchrony in isolated syllables (Lewkowicz, 2000b).

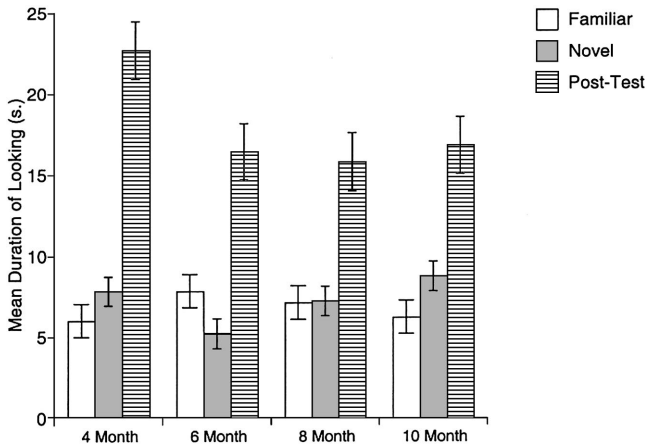


Figure 3. Mean duration of looking in the familiar, the audiovisual desynchrony (novel), and the posttest trials in Experiment 2. Vertical lines depict standard errors of the means.

**Procedure.** Infants at each age were assigned randomly to one of the four groups corresponding to each Actor  $\times$  Rhythm combination. The infant-controlled habituation/test procedure was used; infants first were given the pretest trial cartoon and then were habituated to one of the two rhythms produced by one of the actors. Once they reached the habituation criterion, they then were given the familiar and novel trials (in counter-balanced order across infants in each age group). Following these two test trials, all infants were given the posttest cartoon.

### Results and Discussion

A preliminary, mixed repeated measures ANOVA with follow-up planned comparison tests was first performed to check for possible fatigue effects. Age (4), habituation stimulus (4), and test trial order (2) were the between-subjects factors, and test trial (pretest, familiar, novel, and posttest) was the within-subjects factor. The planned comparison test, contrasting the data from the pre- and posttest trials, indicated that infants actually increased their looking from the beginning to the end of the experimental session,  $F(1, 96) = 3.99, p < .05$ . Moreover, a comparison of the duration of looking in the posttest trial to that in the familiar test trial indicated that infants exhibited significant response recovery to the cartoon presented in the posttest trial,  $F(1, 96) = 160.3, p < .01$ . These findings rule out the possibility that fatigue effects might have accounted for any failures to respond to novelty.

The principal analysis was the same as the preliminary one except that the data from the pre- and posttest trials were not included. Results of this analysis indicated that the test trials factor was not significant but that the Age  $\times$  Trials interaction was statistically significant,  $F(3, 96) = 3.2, p < .05$ . Separate planned comparison tests comparing response in the familiar and novel test trials were performed at each age to determine the source of this interaction. These comparisons indicated that response in the novel test trial was not significantly higher than response in the familiar test trial at 4, 6, or 8 months of age but that there was significant recovery at 10 months of age,  $F(1, 96) = 3.97, p < .05$  (see Figure 3). In addition, as can be seen in Figure 3, there was a familiarity effect at 6 months of age,  $F(1, 96) = 4.13, p < .05$ . To check for possible fatigue effects at each age, I conducted separate planned

comparison tests comparing response in the posttest trial to that in the familiar test trial. Results showed that infants at each age exhibited significant response recovery in the posttest trial ( $ps < .01$ ), meaning that the failure of the younger infants to respond to desynchronization was not due to fatigue.

It should be noted that the identical syllable stimuli were used during the habituation phase in Experiments 1 and 2. The only difference between these two experiments was that the stimulus presented in the novel test trial in Experiment 1 was a different rhythmic pattern, whereas the stimulus presented in Experiment 2 was the familiar, but desynchronized, rhythmic pattern. As a result, it was possible to compare the test trial results in Experiments 1 and 2 directly by way of a mixed repeated measures ANOVA, with experiment (2), habituation rhythm (2), and age (4) as between-subjects factors and test trials (2) as the within-subjects factor. There was a significant Experiment  $\times$  Test Trials interaction,  $F(1, 249) = 6.68, p < .025$ , indicating that the response profile in the familiar and novel test trials differed across the two experiments. In sum, despite the fact that infants of all ages discriminated the rhythmic pattern differences (see Experiment 1), infants younger than 10 months of age did not respond to the intersensory desynchronization of the rhythmic patterns. This developmental pattern of findings is consistent with the hypothesis proposed in the introduction. According to the hypothesis, initially in development infants may not detect desynchronization of a rhythmic pattern, but as they age and their perceptual skills improve, they should become capable of detecting synchrony and rhythmic pattern cues.

### Experiment 3

If rhythmic pattern cues actually prevent infants from detecting intersensory temporal synchrony cues, then removing rhythmic cues should make it possible for the younger infants to perceive desynchronization. I put this possibility to empirical test in Experiment 3 by testing 4-, 6-, and 8-month-old infants' response to syllabic A-V desynchrony in the absence of rhythmic pattern cues. Only one previous study has investigated infants' perception of intersensory temporal synchrony relations inherent in an isolated syllable (Lewkowicz, 2000b). Results indicated that 6- and 8-month-old infants responded to a temporal desynchrony of 666 ms but that 4-month-old infants did not. Because perception of desynchrony was tested only with a single syllable (/ba/) in this earlier study, a secondary aim in the current experiment was to extend the previous findings by testing infants' responsiveness to more than one type of syllable. To do so, and at the same time to provide greater generalizability of findings, I tested infants with one of two distinct types of syllables. They differed in terms of their voicing, place of articulation, and manner of articulation. Infants were habituated to an intermodally synchronous syllable that was uttered repeatedly in a nonrhythmic fashion. Then they were given a novel test trial in which the auditory and visual attributes of the syllable were desynchronized by 633 ms.

### Method

**Participants.** Separate groups of 16 infants each were tested at 4, 6, and 8 months of age. The 4-month-old group of infants (mean age = 18.8 weeks,  $SD = 1.2$  weeks) consisted of 7 boys and 9 girls, the 6-month-old group (mean age = 27.6 weeks,  $SD = 1.2$  weeks) consisted of 9 boys and 7 girls, and the 8-month-old group (mean age = 36.5 weeks,  $SD = 1.0$  week)

consisted of 8 boys and 8 girls. An additional 5 infants were tested but did not contribute data because of fussiness or failure to reach the habituation criterion within 20 trials.

**Apparatus and stimuli.** The Mac IICI system was used for this experiment. The syllables presented during the habituation phase were either /ba/ or /sha/. Two different female actors were filmed as they recited each of the syllables. The films portrayed each actor's head such that the top and bottom of her head corresponded to the top and bottom edges of the stimulus display monitor. Each actor was filmed while she uttered the syllable repeatedly in a relatively irregular fashion with an average inter-syllable interval of 2 s. The manner in which the stimulus materials for this experiment were made differed from that used for making the stimulus materials for Experiments 1 and 2. Thus, for the habituation phase of this experiment, one continuous video recording was made in which the actor was depicted repeatedly uttering the same syllable. During filming, each actor was encouraged to smile and speak in a lively manner, meaning that she was encouraged to vary the pitch and loudness of her voice as she uttered the syllables. In terms of loudness, the SPL of the syllables ranged between 67 and 74 dB (A scale, re .0002 dynes/cm<sup>2</sup>) for the syllables uttered by one of the actors and between 63 and 73 dB for the syllables uttered by the other actor.

For each Actor  $\times$  Syllable combination, a novel test trial movie was made in which the audible and visible components were desynchronized. To accomplish this, I re-recorded each of the four habituation recordings on videotape, but this time I desynchronized the audio portion of the recording with respect to the video portion by having the audio portion precede the video portion by 19 video frames, corresponding to a desynchrony of 633 ms.

**Procedure and design.** As in the prior two experiments, an infant-controlled habituation/test procedure was used. Infants at each age were assigned to one of four habituation groups corresponding to each of the Actor  $\times$  Syllable combinations. The experiment began with the pretest trial (the cartoon), continued with the habituation phase, and ended with the test phase. During the test phase, infants were given a familiar test trial (in which the familiar and synchronous syllable was presented) and a novel test trial, presented in counterbalanced order across infants in each age group. The last trial for all infants was the posttest trial (the cartoon).

## Results and Discussion

The test trial results from this experiment can be seen in Figure 4. A mixed repeated measures ANOVA was first used to analyze the data for possible fatigue effects; age (3), habituation stimulus (4), and test order (2) were the between-subjects factors, and test trial (4) was the within-subjects factor. The analysis yielded a significant trials effect,  $F(3, 72) = 38.7, p < .01$ . A planned comparison test, contrasting the data from the pre- and posttest trials, indicated that looking duration did not differ from the beginning to the end of the experimental session. In addition, a comparison of looking in the posttest trial with looking in the familiar test trial indicated that infants showed significant response recovery to the cartoon presented in the posttest trial,  $F(1, 24) = 79.6, p < .01$ . To determine whether infants detected the audiovisual desynchrony, I repeated the repeated measures ANOVA but this time without the data from the pre- and posttest trials. This analysis yielded a significant trials effect,  $F(1, 24) = 7.02, p < .025$ , and no other effects, indicating that infants detected the desynchrony.

As predicted, in the absence of rhythmic pattern cues, 4- to 8-month-old infants perceived the A-V desynchronization. Moreover, their ability to do so was not affected by the specific nature

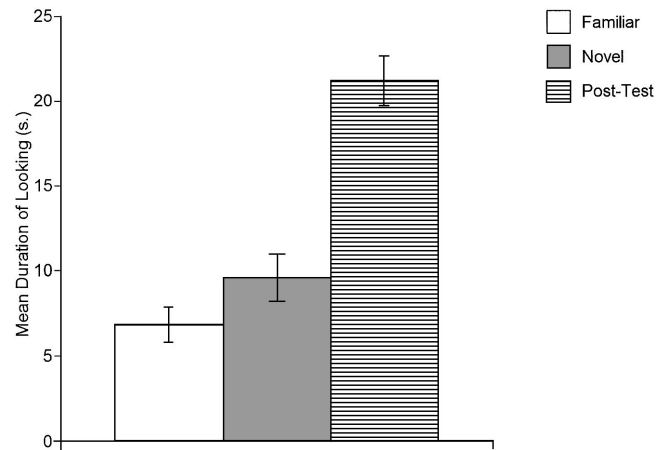


Figure 4. Mean duration of looking in the familiar, the audiovisual desynchrony (novel), and the posttest trials in Experiment 3. Vertical lines depict standard errors of the means.

of the syllable. That is, neither the specific type of syllable nor the person uttering it had an effect on responsiveness. This finding is particularly impressive for two reasons. First, whereas the /ba/ syllable provides relatively obvious bilabial cues that signal the beginning of the syllable, the /sha/ syllable does not. Nonetheless, infants detected the desynchrony equally well, which shows that infants' ability to perceive speech-based audiovisual temporal synchrony relations is relatively robust. Second, statistically reliable results were obtained in the current experiment with half the number of infants tested in Experiment 2. This is clear evidence that infants younger than 10 months of age find the detection of intersensory synchrony relations that are not embedded in a rhythmic pattern easier and/or that rhythmic pattern cues have greater perceptual salience for younger infants than do intersensory temporal synchrony cues.

It should be noted that the syllables presented in Experiment 2 were not uttered in an exaggerated manner, whereas the syllables presented in the current experiment were. Although this raises the possibility that the current results might reflect an effect of manner of speech, this is unlikely in view of the results from an earlier study of mine (Lewkowicz, 2000b). In that study, I investigated the effects of manner of speech on infants' detection of syllabic desynchrony and found that infants detected desynchrony regardless of manner of speech. On the basis of this finding, it can be concluded that the results obtained in the current experiment reflect infants' basic ability to detect speech-based A-V temporal synchrony relations.

## General Discussion

Intersensory temporal synchrony is arguably the simplest and, perhaps for that reason, the developmentally earliest basis for the perception of a multimodally unified world (Lewkowicz, 2000a). To date, studies have assessed whether synchrony can facilitate intersensory perception by itself and/or together with other stimulus features. No studies to date, however, have investigated whether synchrony per se is a discriminable feature in those cases in which it facilitates perception of other features across sensory

modalities. Theoretically, this issue is important because the world is nearly always specified by concurrent stimulus features, and thus, defining the relationship between these various features is important for understanding perception. In the current study, I addressed this question by studying infants' learning and discrimination of bimodally specified and synchronous rhythmic patterns. Because synchrony and rhythm are two temporal cues that are at the opposite ends of the temporal complexity spectrum, infants' response to bimodally specified and synchronous rhythms is likely to yield particularly interesting results *vis-à-vis* the relationship between synchrony and rhythmic pattern cues in learning and discrimination.

Consistent with predictions, the results show that early in development, infants respond only to rhythmic pattern cues but that later they respond to rhythmic pattern cues as well as to temporal intersensory synchrony cues. In summary, Experiment 1 showed that infants between 4 and 10 months of age can discriminate between audiovisual rhythms and that this ability depends neither on the specific temporal pattern that they initially learned nor on whether the pattern consisted of a rhythmically uttered syllable or a rhythmically moving/sounding object. Experiment 2 showed that when the intersensory temporal synchrony of a rhythmically uttered syllable was disrupted, 4- to 8-month-old infants did not respond to this desynchronization but 10-month-old infants did. Finally, Experiment 3 confirmed the greater attention-getting value of rhythmic pattern cues at the earlier ages by showing that 4- to 8-month-old infants did respond to A-V desynchronization of a nonrhythmically presented audiovisual syllable.

It should be noted that stimulus presentation was twice as slow in Experiment 3 as in Experiment 2. As a result, the failure of infants in Experiment 2 to detect the A-V desynchrony and the success of infants in Experiment 3 to do so might have been due to the overall difference in the rate of stimulus presentation rather than to the presence or absence of rhythmic pattern cues. Results from previous studies that investigated detection of A-V desynchrony at different rates of stimulus presentation suggest that this is unlikely. In one study (Lewkowicz, 1992b), I investigated desynchrony detection by presenting a bouncing visual stimulus together with an impact sound, either at a stimulus presentation rate of 0.325 Hz or at a presentation rate of 0.730 Hz. In another study (Lewkowicz, 1996), I presented the same bouncing stimulus and impact sound but this time at a stimulus presentation rate of 0.42 Hz. Regardless of rate of presentation, infants in all three experiments detected desynchronization after being habituated to a synchronous audiovisual event. It should also be noted that the stimulus presentation rate in the present Experiment 2 (1.05 Hz) was very close to the rate of 0.73 Hz in one of my previous experiments. Because a rate of 0.73 Hz yielded successful discrimination, it is unlikely that the failure of infants in Experiment 2 to respond to the desynchronization was because the stimulus presentation rate was too high.

Several aspects of the current findings are noteworthy. First, the current study provides evidence of successful and robust perception of bimodally specified rhythmic patterns across the broadest developmental span so far examined. Prior studies (Bahrick & Lickliter, 2000; Pickens & Bahrick, 1995, 1997) examined perception of audiovisual rhythms only at 5 and 7 months of age. Second, when compared with previous studies of infants' percep-

tion of rhythmic patterns, the current study provides clear evidence that bimodal specification makes rhythmic events more perceptually salient. Specifically, Morrongiello (1984) found that it was not until 12 months of age that infants could detect changes in the relative order of interelement intervals making up different auditory rhythmic patterns. In contrast, the current study showed that infants as young as 4 months of age can detect such changes. The current finding that a bimodally specified event has greater perceptual salience is consistent with a large body of similar evidence from studies that used a wide range of experimental tasks tapping attentional, discriminative, learning, and communication skills in different species and at different stages of development (Lewkowicz, 2002; Partan & Marler, 1999; Rowe, 1999). Finally, perhaps the most interesting and novel finding from the current study is that younger infants responded only to rhythmic pattern cues, whereas older infants responded to rhythmic pattern and synchrony cues. At first blush, this finding appears to be counterintuitive because, as noted earlier, there is substantial evidence that infants as young as 2 months of age can perceive A-V synchrony relations and that slightly older infants can make synchrony-based intersensory matches (Lewkowicz, 2000a). As the hypothesis that initially motivated this research suggests, however, it makes sense from a global processing point of view that rhythmic pattern cues might override the perception of A-V synchrony cues. In fact, the kind of hierarchical relationship found here is consistent with other findings. For example, infants do not make intersensory matches on the basis of joint temporal synchrony and rate cues, but they do make such matches on the basis of synchrony cues alone (Lewkowicz, 1992a, 1994). Likewise, infants are sensitive to the temporal distributional properties of speech (Saffran, Aslin, & Newport, 1996), but they do not respond to them when such cues are put in competition with stress or coarticulation segmentation cues (Johnson & Jusczyk, 2001).

One of the interesting questions that the current study and its findings raise is whether infants were perceiving the intersensory unity of the rhythmic pattern information and whether this, in part, accounted for their responsiveness to rhythmic pattern cues in early development. There are two ways to perceive intersensory unity in the case of a bimodal rhythmic pattern. Either one can perceive the two sensory streams specifying the rhythmic structure in an additive fashion, or one can perceive them in a truly integrated manner as a single, amodally invariant pattern. The experimental method used in the current study does not permit an assessment of these two possibilities. One method that does permit such an assessment is a cross-modal transfer procedure. It explicitly tests infants' ability to perceive amodal invariance in the absence of intersensory temporal synchrony cues by first allowing infants to learn a particular rhythmic pattern in one modality and then testing them for recognition of that particular pattern in a second modality. In the absence of data on cross-modal transfer, the most conservative interpretation of the greater salience of bimodal specification is that it was due primarily to the synchrony of the audible and visible components. The synchrony made the auditory and visual attributes of the rhythmic pattern contribute to intersensory redundancy, but that redundancy is the result of additive, rather than intermodally integrative, processes (Lewkowicz & Kraebel, *in press*). This interpretation is consistent with my model (Lewkowicz, 2000a) of the development of intersensory perception of equivalence. The model predicts that the ability to

perceive the intersensory equivalence of rhythmic pattern information does not emerge until quite late in infancy. As a result, the most reasonable interpretation of the greater salience of bimodal specification is that early in development the redundancy benefit is due to additive effects but later in development it also may be due to intersensory integration.

The finding that the 10-month-old infants responded to both the synchrony and the rhythmic pattern cues raises questions about possible developmental changes in underlying perceptual mechanisms. If, as suggested by my model (Lewkowicz, 2000a), infants do not perceive the intersensory equivalence of rhythmic pattern information until late in infancy, and if younger infants process the auditory and visual information as two separate streams, then the younger infants are faced with a heavier processing load. This is because processing two streams of information in an additive fashion most likely imposes a greater processing load than does processing two streams as a single integrated stream. If this is correct, then once infants reach an age at which they can perceive the intersensory equivalence of rhythmic pattern information, processing load is presumably reduced and the perception of A-V synchrony relations then becomes possible as well. This developmental scenario suggests that the perception of intersensory temporal synchrony relations in a bimodal learning and discrimination task depends on the concurrent presence or absence of other temporal cues. Early in development (i.e., prior to 10 months of age), perception of A-V synchrony relations is essentially blocked by higher order temporal pattern cues, but later in development it is not because of newly emerging intersensory perception skills that permit intersensory integration of higher level temporal pattern information.

Although the information-processing interpretation offered to account for the current findings is somewhat speculative, it is based on a firm empirical and theoretical footing and is eminently testable. The first step that future studies might take in testing the validity of the current interpretation is to investigate the cross-modal transfer of rhythm across a wide age span in infancy to determine when this ability first emerges. Once the developmental emergence of the intersensory perception of rhythmic pattern equivalence is empirically investigated, it will be possible to better characterize the mechanisms that enable infants to take advantage of intersensory redundancy when learning and discriminating bimodally specified objects and events.

## References

- Bahrick, L. E. (1988). Intermodal learning in infancy: Learning on the basis of two kinds of invariant relations in audible and visible events. *Child Development, 59*, 197–209.
- Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology, 36*, 190–201.
- Baldwin, D. A., & Baird, J. A. (2001). Discerning intentions in dynamic human action. *Trends in Cognitive Sciences, 5*, 171–178.
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Bushara, K. O., Grafman, J., & Hallett, M. (2001). Neural correlates of auditory–visual stimulus onset asynchrony detection. *Journal of Neuroscience, 21*, 300–304.
- Chang, H. W., & Trehub, S. E. (1977). Auditory processing of relational information by young infants. *Journal of Experimental Child Psychology, 24*, 324–331.
- Demany, L., McKenzie, B., & Vurpillot, E. (1977). Rhythm perception in early infancy. *Nature, 266*, 718–719.
- Dixon, N. F., & Spitz, L. T. (1980). The detection of auditory visual desynchrony. *Perception, 9*, 719–721.
- Dodd, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology, 11*, 478–484.
- Fraisse, P. (1982a). The adaptation of the child to time. In W. J. Friedman (Ed.), *The developmental psychology of time* (pp. 113–140). New York: Academic Press.
- Fraisse, P. (1982b). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149–180). New York: Academic Press.
- Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44*, 548–567.
- Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin, 126*, 159–179.
- Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior: The Hixon symposium* (pp. 123–147). New York: Wiley.
- Lewkowicz, D. J. (1986). Developmental changes in infants' bisensory response to synchronous durations. *Infant Behavior & Development, 9*, 335–353.
- Lewkowicz, D. J. (1989). The role of temporal factors in infant behavior and development. In I. Levin & D. Zakay (Eds.), *Time and human cognition: A life-span perspective* (pp. 9–62). Amsterdam: North-Holland.
- Lewkowicz, D. J. (1992a). Infants' response to temporally based intersensory equivalence: The effect of synchronous sounds on visual preferences for moving stimuli. *Infant Behavior & Development, 15*, 297–324.
- Lewkowicz, D. J. (1992b). Infants' responsiveness to the auditory and visual attributes of a sounding/moving stimulus. *Perception & Psychophysics, 52*, 519–528.
- Lewkowicz, D. J. (1994). Limitations on infants' response to rate-based auditory–visual relations. *Developmental Psychology, 30*, 880–892.
- Lewkowicz, D. J. (1996). Perception of auditory–visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1094–1106.
- Lewkowicz, D. J. (2000a). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin, 126*, 281–308.
- Lewkowicz, D. J. (2000b). Infants' perception of the audible, visible and bimodal attributes of multimodal syllables. *Child Development, 71*, 1241–1257.
- Lewkowicz, D. J. (2002). Heterogeneity and heterochrony in the development of intersensory perception. *Cognitive Brain Research, 14*, 41–63.
- Lewkowicz, D. J., & Kraebel, K. (in press). The value of multimodal redundancy in the development of intersensory perception. In G. Calvert, C. Spence, & B. Stein (Eds.), *Handbook of multisensory processing*. Cambridge, MA: MIT Press.
- Martin, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review, 79*, 487–509.
- Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.
- McGrath, M., & Summerfield, Q. (1985). Intermodal timing relations and audio–visual speech recognition by normal-hearing adults. *Journal of the Acoustical Society of America, 77*, 678–685.
- Mendelson, M. J. (1986). Perception of the temporal pattern of motion in infancy. *Infant Behavior & Development, 9*, 231–243.

- Morrongiello, B. A. (1984). Auditory temporal pattern perception in 6- and 12-month-old infants. *Developmental Psychology, 20*, 441–448.
- Partan, S., & Marler, P. (1999). Communication goes multimodal. *Science, 283*, 1272–1273.
- Pickens, J., & Bahrack, L. E. (1995). Infants' discrimination of bimodal events on the basis of rhythm and tempo. *British Journal of Developmental Psychology, 13*, 223–236.
- Pickens, J., & Bahrack, L. E. (1997). Do infants perceive invariant tempo and rhythm in auditory–visual events? *Infant Behavior & Development, 20*, 349–357.
- Pomerantz, J. R., & Lockhead, G. R. (1991). Perception of structure: An overview. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure: Essays in honor of Wendell R. Garner* (pp. 1–20). Washington, DC: American Psychological Association.
- Radeau, M. (1994). Auditory–visual spatial interaction and modularity. *Cahiers de Psychologie, 13*, 3–51.
- Rowe, C. (1999). Receiver psychology and the evolution of multicomponent signals. *Animal Behaviour, 58*, 921–931.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.
- Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology, 15*, 626–636.
- Thorpe, L. A., & Trehub, S. E. (1989). Duration illusion and auditory grouping in infancy. *Developmental Psychology, 25*, 122–127.
- Thorpe, L. A., Trehub, S. E., Morrongiello, B. A., & Bull, D. (1988). Perceptual grouping by infants and preschool children. *Developmental Psychology, 24*, 484–491.
- Trehub, S. E., Schneider, B. A., & Henderson, J. L. (1995). Gap detection in infants, children, and adults. *Journal of the Acoustical Society of America, 98*, 2532–2541.
- Werner, L. A., Marean, G. C., Halpin, C. F., Spetner, N. B., & Gillenwater, J. M. (1992). Infant auditory temporal acuity: Gap detection. *Child Development, 63*, 260–272.
- Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127*, 3–21.

## Appendix

### Terms Relevant to a Discussion of Intersensory Perception and Their Conceptual and Operational Definitions

| Term                     | Conceptual Definition  | Operational Definition Examples  |
|--------------------------|--|--|
| Intersensory equivalence | The perceptual effect of a class of stimulus properties that are amodal by virtue of the fact that they can be specified equally well in different sensory modalities. Examples of such stimulus properties include intensity, shape, texture, duration, rate, rhythm, and location. The perceptual effect is that of equivalence regardless of modality of specification. | In vision and audition, when watching and listening to a talker, one can see that the duration, tempo, and rhythm of mouth opening and closing are the same as the duration, tempo, and rhythm of voice modulation. In touch and vision, one can tell that a cube is the same whether it is seen or felt.                    |
| Intersensory matching    | The process by which a subject perceives an equivalent property in different sensory modalities. It can be accomplished either when an object or event is specified concurrently by an amodal property or when the amodal property is made available first in one sensory modality and later in a second modality (the latter is also known as cross-modal matching).      | In vision and audition, intersensory matching is successful when a given visual rhythm is perceived as being the same when it is heard. In vision and touch, intersensory matching is successful when one can see a given object's shape or texture and then recognize these properties as the same when feeling the object. |
| Intersensory integration | The process by which sensory inputs from different modalities are combined into unified and often meaningful percepts. This process can rely on the perception of intersensory equivalence, intersensory/cross-modal matching, and/or on the association of unique modality-specific properties (e.g., color, pitch, smell).   | While watching and listening to a talker, one can perceive the amodal properties of the talker's face and voice and at the same time associate the talker's hair color, pitch of voice, and smell.   |
| Intersensory redundancy  | The condition in which an object or event may be specified by some combination of amodal and modality-specific attributes. Usually, such stimulus properties are available at the same time and in the same spatial location.  | A talker can be specified by a variety of amodal invariants such as duration, tempo, and rhythm as well as by pitch, color, and smell.   |

Received May 15, 2002  
 Revision received February 28, 2003  
 Accepted March 7, 2003 ■