Intermodal Learning in Infancy: Learning on the Basis of Two Kinds of Invariant Relations in Audible and Visible Events

Lorraine E. Bahrick
Florida International University

BAHRICK, LORRAINE E. *Intermodal Learning in Infancy: Learning on the Basis of Two Kinds of Invariant Relations in Audible and Visible Events*. CHILD DEVELOPMENT, 1988, **59**, 197–209. In this research, the development of intermodal perception in infancy was examined by using a new method, the intermodal learning method. 3-month-old infants were given the opportunity to learn a relation between 2 single film and soundtrack pairs through a 2-min familiarization period under 1 of 4 conditions. Films of naturalistic events were accompanied by a soundtrack that was (1) appropriate to the composition of the object and synchronous with its motions, (2) appropriate and nonsynchronous, (3) inappropriate and synchronous, or (4) inappropriate and nonsynchronous. A group of control subjects was familiarized with irrelevant films and soundtracks. Then all subjects were tested in a 2-choice intermodal preference test to determine under which familiarization conditions intermodal learning had occurred. Results indicated that only subjects who had been familiarized with appropriate and synchronous film and soundtrack pairs showed evidence of intermodal learning as compared with the performance of control subjects. Intermodal learning occurred on the basis of 2 kinds of invariant audio-visual relations, temporal synchrony, and temporal microstructure specifying the composition of the object. Intermodal learning did not occur through association on the basis of co-occurrence, nor did it occur when any incongruent audio-visual structure was present. These findings support an invariant-detection view of the development of intermodal perception.

In recent years, a great deal of research has focused on the development of intermodal perception in infancy. This research has been motivated by two predominant theoretical positions. The traditional integration-association view (Birch & Lefford, 1963, 1967; Blank & Bridger, 1964; Bryant, 1974) holds that infants must learn to put together or integrate sensory impressions from separate input modalities through a gradual process of association. Until this integration is achieved, the infant can perceive only unrelated patterns of light and sound from a single multimodal event. A second position is the invariant-detection view of perceptual development (Gibson, 1969). According to this view, the senses are unified at birth, and the infant is endowed with the capacity for detecting invariants in stimulation, relations that remain constant across stimulus transformations. Multimodal events often provide amodal invariant relations that unite the stimulation from different sensory modalities. Amodal information is not specific to a particular sensory modality; rather, the same information can be picked up through two or more modalities. For example, the audible and visible stimulation from a bouncing ball is united by an amodal, invariant temporal structure. One can detect the same rhythm and tempo by watching the ball's motion, or by listening to its impact sounds. By detecting this invariant structure, the infant can perceive a unitary audible and visible event from the beginning, making the process of association unnecessary. The developmental task of the child, according to this view, is to learn to differentiate and abstract progressively finer levels of invariant structure.

Proponents of the invariant-detection view have focused on delineating the nature of intermodal invariants that infants are capable of detecting. Two basic methods for assessing these capabilities have predominated—the cross-modal transfer design, and Spelke's (1976) intermodal preference method. A third method, the conflict procedure, where infants' reactions to conflicting visual and haptic or auditory information are assessed, has produced evidence of intermodal functioning in only some studies

This research was supported by National Institute of Child Health and Human Development grants 1 R01 HD15725 and 5 R23 HD18766. Special thanks are extended to Norene Huber, Betty Kaufman, Deborah Kennedy, and Nancy Roscelli for their assistance in the collection and analysis of data, and to John S. Watson for his helpful suggestions. Requests for reprints should be sent to Lorraine Bahrick, Department of Psychology, Florida International University, Miami, FL 33199.

[*Child Development*, 1988, **59**, 197–209. 0009-3920/88/5901-0001$01.00]

(Bower, Broughton, & Moore, 1970a, 1970b; Bushnell, 1981; Lasky, 1977).

In the cross-modal transfer design, infants are familiarized with an object or event through one sensory modality, and their behavior toward that object or event presented to a different sensory modality is then assessed. Researchers using this approach have demonstrated that young infants detect invariant relations across visual and tactile modalities specifying the shape, texture, and substance of objects (Gibson & Walker, 1984; Gottfried, Rose, & Bridger, 1977; Meltzoff & Borton, 1979; Ruff & Kohler, 1978), as well as the invariant temporal pattern or rhythmic structure across visual and acoustic presentations (Allen, Walker, Symonds, & Marcell, 1977).

In the intermodal preference method, subjects are typically presented with two filmed events side by side, along with a soundtrack to one of them coming from a centrally located speaker. The effects of the acoustic stimulation on their visual exploration of the two events is then assessed. Using this method, it has been demonstrated that by the age of 4½ months, infants detect temporal synchrony relations (Bahrick, 1983; Dodd, 1979; Spelke, 1979, 1981; Spelke, Born, & Chu, 1983; Spelke & Cortelyou, 1980; Walker, 1982), the common tempo of action (Spelke, 1979), and the common rhythmic structure (Mendelson & Ferland, 1982) uniting the visual and acoustic stimulation from natural filmed events. They also detect the common audio-visual temporal microstructure specifying the substance and the composition of an object colliding against a surface (Bahrick, 1983, 1987) prior to the age of 6 months.

These studies have firmly established that infants in the first 6 months of life are capable of detecting invariant information uniting the stimulation from different sensory modalities. This finding in itself, however, reveals little about the origins of these capabilities, or how they lead to the acquisition of intermodal knowledge about objects and events in the environment. A more promising approach for exploring the consequences of these capabilities has been that of examining the process of intermodal learning itself, and assessing whether and under what conditions the infant's detection of invariants leads to knowledge about visible properties of acoustically detected events.

Spelke (1978) extended her intermodal preference method to address these questions. Subjects who were given the opportunity to abstract the invariant relations uniting one of two moving objects with a soundtrack in the intermodal preference method were then given a visual "search" test to assess learning. Using this method, it has been demonstrated that infants learn about the visible properties of acoustically detected events through abstracting amodal invariants such as temporal synchrony (Bahrick, in press; Spelke, 1979, 1981), the common tempo of action (Spelke, 1979), and the common temporal microstructure specifying the substance and the composition of an object colliding against a surface (Bahrick, 1983, in press).

A second approach for investigating intermodal learning (e.g., Lawson, 1980; Lyons-Ruth, 1977) has been to attempt to teach the infant a specific object-sound relation by presenting a single object paired with a single type of sound for familiarization. Then, fixation to both the familiar object and a novel one in the presence of the familiar sound is assessed. In contrast with the intermodal preference and search method, learning on the basis of association is possible here since a single object and sound co-occur during familiarization. Using this method, Lawson (1980) found that infants learned about object-sound relations on the basis of temporal synchrony when the object and sound were colocated, but not when they were widely displaced. Using an analogous method with visual and tactile familiarization, Bushnell (1986) found limited evidence of intermodal learning for objects differing in shape and texture.

By now it is quite well established that young infants possess a variety of interesting intermodal capabilities and that some of these capabilities provide the basis for intermodal learning about the visible properties of acoustically detected events. Although these findings are clearly consistent with the invariant-detection view of perceptual development, the origin of these intermodal capabilities continues as a topic of much speculation. How do these capabilities emerge and develop? Which invariant relations are detected first, and how does detection of one invariant lead to detection of others? Though studies to date have provided an important data base concerning the nature of intermodal relations that infants are capable of detecting and which of these provide the basis for learning about specific object-sound relations, their methodological approach has been less appropriate for answering questions about the origins and developmental course of these capabilities. The learning methods described above have almost exclusively been used to assess learning on the basis of already developed intermodal capabilities. Furthermore, many of the intermodal preference and cross-modal transfer studies have been used to assess infants' detection of only a single type of intermodal invariant, for a single stimulus contrast, or to test infants of only one particular age.

For the purposes of the present research, a different method for assessing the nature of intermodal learning was developed, specifically to test younger infants who initially show no knowledge of the audio-visual invariants in question. It is called the intermodal learning method and entails familiarizing the infant with two audible and visible events, one at a time, and then assessing the conditions under which intermodal learning took place. The intermodal learning method combines some of the most attractive features of the prior learning methods. The learning process can be examined directly by giving subjects the opportunity to learn the relation between a visible object and its sound during familiarization, similar to the methods used by Lawson (1980) and Lyons-Ruth (1977). Then, one can assess under which familiarization conditions learning took place in a manner similar to Spelke's (1976) intermodal preference method. This method has several advantages over prior methods when it is used, as in the present research, with younger infants who do not yet show evidence of spontaneously detecting the audio-visual invariants in question. First, by examining the learning process directly in younger, less experienced infants, it brings us closer to the origins of the learning process. Second, one can assess whether certain object-sound relations are learned more readily than others, to determine the kinds of constraints that are inherent in the learning process early on. Third, given that infants show no initial capability for detecting the object-sound relations, one can use natural, ecological audio-visual events, since there is then no need to control for prior learning by presenting arbitrary object-sound combinations as in prior learning studies (Lawson, 1980; Lyons-Ruth, 1977). Further, the use of natural events enables one to investigate the infant's knowledge about the kinds of sounds different visually specified events naturally produce. And finally, by choosing natural events that make two or more types of invariant relations available, one can examine how multiple invariants interact in directing the process of intermodal learning prior to the age where these intermodal capabilities are evident in a two-choice intermodal preference test. The present research incorporates all of these features into a single study where infants' learning on the basis of two types of bimodal temporal invariants is explored.

A recent study by Bahrick (in press) provided a point of departure for the research reported here. Infants' sensitivity to two levels of invariant temporal structure uniting the audible and visible stimulation from natural events was assessed using an intermodal preference method. Films depicting a single large marble (single object) and a group of smaller marbles (compound object) colliding against a surface in an erratic pattern were presented to infants between the ages of 3 and 7½ months. These stimulus events were characterized by two levels of nested invariant temporal structure. A global temporal synchrony relation (temporal macrostructure) united the sights and sounds of the objects' impacts, while an embedded level of audio-visual temporal structure (temporal microstructure) characterized the nature of each synchronous impact motion and sound. In this case, the temporal microstructure specified the composition of the moving object: whether it was composed of a single, solid substance, or a group of smaller units. Results of the study revealed that by the age of 6 months, but not at 3 or 4½ months, infants were capable of detecting the temporal microstructure specifying object composition. They looked predominantly to the film whose natural soundtrack was played, even though the motions of both the single and compound object were synchronized with the soundtrack. However, when the motions of neither object were synchronized with the soundtrack, the temporal asynchrony disrupted their detection of the intermodal relations. These findings demonstrated that while 6-month-olds can detect the relation between a film and soundtrack on the basis of either type of temporal invariant, they perceive no relation between audible and visible stimulation when incongruent temporal structure is detected.

The present research was conducted to explore the developmental basis for this finding. The intermodal learning method was used to determine under what conditions 3-month-old infants would learn to perceive a relation between the audible and visible stimulation from these events. The same stimulus films were used as in the study described above. Infants were given the opportunity to learn the relation between a single film and

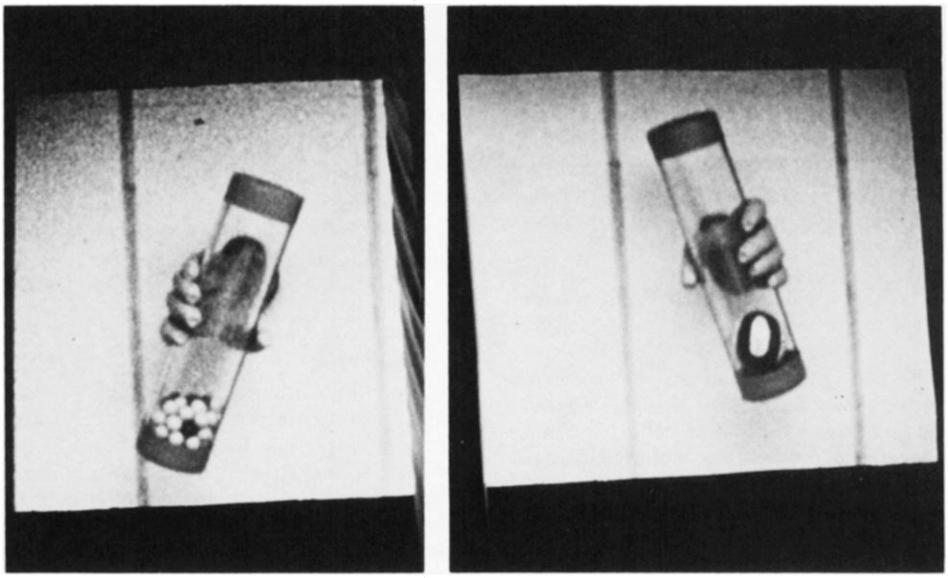

FIG. 1.—Photograph of the stimulus events

accompanying soundtrack during a 2-min familiarization period. They participated in one of four kinds of familiarization conditions in which the two levels of temporal structure provided either congruent or incongruent audio-visual relations. The soundtracks were presented either in or out of synchrony with the motions of the filmed objects, and the sounds were either appropriate or inappropriate to the composition of the objects. Following familiarization, infants in all conditions were tested in a two-choice intermodal learning test to determine under which familiarization conditions learning had occurred. A group of age-matched control subjects who were not exposed to the stimulus events during the familiarization period showed no preference for the acoustically specified films in the two-choice test. Therefore, this test provided a measure of intermodal learning for subjects in the other conditions. Thus, infants' detection of two kinds of audio-visual invariants in a single event was examined, and how these two levels of invariant structure interact in directing intermodal learning was assessed.

Method

Subjects.—One hundred 3-month-old infants whose mean age was 98.5 days (SD = 4.13 days) participated in the experiment. Twenty-four additional subjects began testing but were unable to complete the procedure. Seventeen were rejected because of excessive fussiness, two fell asleep, and five were rejected because of equipment failure or experimenter error. Subjects were recruited through the local birth records, and all were healthy, with no complications upon delivery.

Stimulus events.—Super-8 sound and color films of two kinds of objects moving in an erratic pattern were used (see Fig. 1). These were the same stimulus films used by Bahrick (in press) and were designed such that two levels of nested invariant temporal structure characterized their motions. Temporal macrostructure, or synchrony, united the sights and sounds of the objects' impacts. A nested level of temporal microstructure, apparent at each impact, specified whether the object was composed of a single, solid substance or an aggregate of smaller units. One film (single object) depicted a single, large, blue and yellow marble, with a diameter of 6.2 cm, inside a transparent Plexiglas cylinder. The other film (compound object) depicted a group of 43 small blue and yellow marbles, also inside a transparent Plexiglas cylinder. There were 15 blue and 28 yellow marbles, each 1.5 cm in diameter. All marbles were solid and made of glass. The cylinders containing both kinds of objects were identical and were 27.5 cm long and had a diameter of 7 cm. They were abruptly rotated in an arc-shaped trajectory so that each sudden impact of the marble(s) against the lower surface of the cylinder could be both seen and heard. The natural soundtracks to each event were recorded live. The single object produced a single, discrete impact sound, while the compound object produced a more prolonged ensemble of sounds. Both events had erratic and unpredictable temporal patterns yet displayed similar rates of motion across the 2-min filmed episodes (approximately 44 impacts per minute). This ensured that infants could not discriminate between the two films on the basis of temporal pattern or tempo of action.

Several filmed versions of the two objects were made to fulfill the various conditions of the experiment. All films were made with a Eumig 128 XL super-8 sound movie camera. For the familiarization period of the experiment, films of each object accompanied by two different kinds of soundtracks were made. One type of film depicted the object moving in synchrony with its natural impact sounds, while the other depicted the object moving in synchrony with the natural impact sounds of the other object. The sound-inappropriate films were made with the use of a soundproof box that was constructed with a Plexiglas front. Both types of objects were moved in synchrony with one another, but only the silent one was filmed through the front of the soundproof box, while the impact sounds of the other object were simultaneously recorded onto the film strip. In addition, the sounds of these objects were tape-recorded, with one soundtrack on each channel, so that these soundtracks could also be played out of synchrony with the motions of the filmed objects during the familiarization period of the experiment.

For the intermodal learning test, films were made of the two kinds of objects moving in synchrony with one another, side by side. In one version, the motions of both objects were synchronized with the natural impact sounds of the single marble, and in the other they were synchronized with the natural impact sounds of the group of smaller marbles. This was accomplished by placing the silent object in the soundproof box while both objects were moved in synchrony with one another and filmed together.

Apparatus.—Infants were seated in an infant seat in a three-sided wooden booth. They faced a split projection screen about 64 cm away. The film images were about 34 × 34 cm and were separated by a distance of 16 cm. A strip of colored Christmas tree lights was positioned between the two screens, and a remote-control mechanical toy dog was centered above the screen and encircled with lights. This display was used to attract the infant's attention to center screen between trials. Two Eumig S926GL super-8 color and sound movie projectors were used to display the films. An Akai M9 two-channel, reel-to-reel tape-recorder was used to play the nonsynchronized soundtracks during two of the familiarization conditions. The soundtracks from both movie projectors, and both channels of the tape-recorder were played through a single speaker centered between the two split screens and about 1½ m away. This ensured that infants could not localize the sound at either the right- or left-hand images. A switch enabled us to transmit one of the four sound outputs to the central speaker and also facilitated a rapid transition from one type of sound output to another.

One or two trained observers monitored the subjects' visual fixations to the two sides of the split screen from apertures located just below the screen. The aperture used by the primary observer was encircled by a cardboard occluder that prevented her from seeing the images projected to the screens even if she turned her head. In addition, both observers were naive with respect to each subject's experimental condition. Observers depressed one button while the subject fixated the right-hand screen and another while the subject fixated the left-hand screen. The fixations were recorded with an eight-channel Rustak strip-chart recorder. The durations of the light display as well as the soundtracks to the filmed events were also recorded by the event recorder.

Procedure.—Three-month-old infants were chosen for the present research, since in a previous series of studies with the same stimulus events, Bahrick (in press, Exp. 1) had found that infants of this age show no evidence of prior knowledge of the audiovisual relations in question. One hundred 3-month-old infants were randomly assigned to one of the four familiarization conditions or a control group. Thus, 20 subjects participated in each of the five conditions. The four familiarization conditions were designed to provide infants with the opportunity to learn a relation between a given type of object and soundtrack. In the four familiarization conditions, single films and soundtracks were presented either in synchrony or out of synchrony with one another, and the films were presented with either the appropriate or inappropriate soundtracks. Thus, the film and soundtrack pairings across the four conditions were (1) appropriate and synchronous (A/S), (2) appropriate and nonsynchronous (A/N), (3) inappropriate and synchronous (I/S), or (4) inappropriate and nonsynchronous (I/N). Therefore, subjects in the A/S condition viewed films of the single marble and group of smaller marbles along with their natural synchronized soundtracks. Those in the A/N condition viewed films of the single marble and group of smaller marbles along with their natural soundtracks, but they were played out of synchrony with the motions of the objects. Subjects in the I/S condition viewed the film of the single marble moving in synchrony with the impact sounds of the group of smaller marbles and the film of the group of small marbles moving in synchrony with the impact sounds of the single large marble. And, finally, those in the I/N condition viewed the film of the single marble along with nonsynchronous impact sounds of the group of small marbles and the film of the group of small marbles along with the nonsynchronous impact sounds of the single marble.

During familiarization, infants received eight 15-sec trials of single film and soundtrack pairs. On four of those trials, the film of the single object was presented to one side of the split screen, and on the other four trials the film of the compound object was presented to the other side of the split screen. Films of the two objects were presented in an alternating pattern so that a particular infant might first view the film of the single object on the right-hand screen along with its designated soundtrack, followed by the film of the compound object on the left-hand screen, along with its designated soundtrack. This continued until eight such trials had been presented. Each intertrial interval was approximately 3 sec long, and during the interval the flashing colored lights and the mechanical toy dog were turned on to attract the infant's attention to center screen. Thus, across the eight trials, each infant received a total 1-min exposure to the film of the single object and a total 1-min exposure to the film of the compound object. The lateral positions of the two films were counterbalanced across subjects in each familiarization condition, and the side on which the alternating pattern began was also counterbalanced across subjects. An observer, naive with respect to the familiarization condition, monitored the subjects' visual fixations to the two films.

Subjects in the control group were familiarized with films of irrelevant events in a manner similar to that described above. These irrelevant events depicted two wooden blocks striking a surface and two water-soaked sponges squashing against one another in an erratic pattern. They had been used as stimulus events in a prior study with 4-month-olds (see Bahrick, 1983). Several factors were taken into account in designing the stimulus-presentation format for subjects in the control group. The purpose of the control group was to control for prior exposure to the test stimuli and to provide a basis of comparison for performance on the intermodal learning test for subjects in the four experimental conditions. It was therefore important that subjects in the control group view films in a manner that was equally appropriate to all four experimental conditions. This eliminated the possibility of presenting films and soundtracks in a manner that duplicated one of the familiarization conditions and not the others (e.g., presentation of only an appropriate and synchronous film and soundtrack pair). To avoid this problem, a presentation format that was identical to none of the familiarization conditions, yet preserved the important parameters of all of them, was used. Control subjects thus received eight 15-sec trials of the two kinds of irrelevant events side by side. During each trial, the soundtrack was synchronous and appropriate to one of the events, and nonsynchronous and inappropriate to the other event. Across the eight trials, the two soundtracks were again presented in an alternating pattern as in the four experimental conditions, so that the sound-specified film might be the one on the right, and then the one on the left, and so forth, for a given infant. The lateral positions of the two irrelevant events and the type of soundtrack presented first were counterbalanced across subjects. All other details of the procedure were identical to those of the other four familiarization conditions. This method succeeded in preserving the eight-trial presentation format, the alternating soundtrack pattern, and the total amount of time infants could spend viewing films during familiarization without favoring one type of film and soundtrack pairing over the other. An observer, who was unaware of the lateral positions of the two films, monitored the subjects' visual fixations.

It should be noted, however, that though subjects in the present control group received a stimulus-presentation format that was comparable to that of the experimental groups, they also were exposed to the test stimuli for a shorter total time since they were not familiarized with them. It is conceivable that a longer exposure to the test stimuli could have enabled even 3-month-olds to manifest some knowledge of the audio-visual relations in question, and therefore a control for this would be appropriate. A previous study mentioned earlier (Bahrick, in press, Exp. 1) provided such a control. Twenty-four 3-month-old infants viewed these two filmed events, along with a soundtrack that was appropriate and synchronous with only one of them, in a typical intermodal preference and search phase. Though they received a total 6-min exposure to the events, they showed no significant looking to the sound specified films (M = .53, preference phase; M = .49, search phase). It was only by the age of 6 months that infants showed significant evidence of detecting a relation between these films and soundtracks on the basis of temporal synchrony and/or sound appropriateness. Thus, even when both kinds of invariant relations united the films and soundtracks of the present events, across a 6-min presentation, 3-month-olds showed no knowledge of the audio-visual relations in question. Since these results were comparable to those of the present control group, only the present results will be used for subsequent analyses.

After participating in one of the five familiarization conditions, the subject was turned around in the infant seat and entertained for a period of approximately 3 min. Then the infant and seat were rotated back into position, and the intermodal learning test began.

The purpose of the intermodal learning test was to assess under which familiarization conditions the infant had learned to perceive a relation between the film and soundtrack pairs presented together. Therefore, all 100 infants participated in an identical intermodal learning test. During the test, films of the single and compound object were presented together, side by side, across eight 15-sec trials. The two objects moved in synchrony with one another and were accompanied by a soundtrack that was synchronous with the motions of both. On four of the trials the natural soundtrack to the single object was played, and on the other four trials the natural soundtrack to the compound object was played. Thus, on a given trial, the infant viewed films of the two objects side by side along with a soundtrack that was synchronous with the motions of both but appropriate to the composition of only one. The intertrial interval was again approximately 3 sec long, and the infant's attention was directed to center screen between trials by the colored lights and mechanical toy dog. The two soundtracks were presented in one of five predetermined orders so that infants could not learn to anticipate the lateral position of the sound-appropriate film across the eight trials. The presentation orders were random with the restrictions that a given soundtrack be played no more than twice in succession and that four presentations of each soundtrack be played across the eight trials. Soundtrack presentation order was counterbalanced so that four infants in each of the five familiarization conditions were randomly assigned to a given presentation order. The lateral positions of the two films during the test were the same as those the infant had received during familiarization, except for control subjects who had not previously seen these films. A trained observer monitored the subject's visual fixations to the two sides of the split screen on each trial. A second observer also monitored the subject's fixations for 29% of the sample for the purposes of calculating interobserver reliability.

Results and Discussion

Familiarization period.—Fixation times for the familiarization periods were derived by calculating the number of seconds the infant spent fixating the single film on each 15-sec trial. These eight fixation times were then averaged to yield a single mean fixation time for each infant. Results indicated that across the four familiarization conditions subjects fixated the films approximately 12.4 sec out of the total possible 15 sec available for each trial, or 83% of the time. These mean fixation times did not differ significantly across the four types of familiarization conditions, $F(3,76) = 1.22$, $p > .05$, according to a one-way analysis of variance. The means for the A/S, A/N, I/S, and I/N conditions were 11.9, 12.6, 12.4, and 12.7, respectively. These results indicate that during the familiarization period, infants did not attend to the films differentially as a function of the type of familiarization they received.

Intermodal learning test.—To test the main hypotheses, results of the intermodal learning test were examined as a function of the type of familiarization infants had previously received. Results of the intermodal learning test were expressed in two nonorthogonal ways. To determine under which familiarization conditions intermodal learning had occurred, the intermodal learning test results of infants in each of the four experimental familiarization conditions were compared against those of the control group. For these analyses, results of the intermodal learning test were expressed as the mean proportion of total fixation time infants in each familiarization condition spent fixating the sound-appropriate film. These mean proportions were de-

TABLE 1

INTERMODAL LEARNING TEST: MEAN PROPORTIONS OF TOTAL LOOKING TIME AND STANDARD DEVIATIONS AS A FUNCTION OF FAMILIARIZATION CONDITION

	FAMILIARIZATION CONDITION				
	Control	Appropriate/ Synchronous	Appropriate/ Nonsynchronous	Inappropriate/ Synchronous	Inappropriate/ Nonsynchronous
A: Looking to the sound-specified film:					
Proportion	.53	.62	.51	.50	.55
SD	.09	.13	.08	.10	.07
B: Looking to the film that had been paired with the soundtrack during familiarization:					
Proportion	...	.62	.51	.50	.45
SD	...	.13	.08	.10	.07

rived by first calculating the proportion of time the infants spent fixating the sound-appropriate film out of the total time they spent fixating both films on each test trial. Proportions for all eight trials were then averaged to yield a single mean proportion for each infant. Mean looking proportions to the sound-appropriate film for subjects in each familiarization condition were then derived from these scores. These proportions are depicted in Table 1A. Note that these proportions are useful for assessing whether or not learning occurred relative to the performance of control subjects, but they do not reflect the direction of the learning effects.

A second method of expressing the results of the intermodal learning test was used to examine questions concerning the nature of the intermodal learning observed as a function of the type of familiarization condition in which the subject had participated. For these analyses, the mean proportion of total looking time infants spent fixating the film that had previously been paired with the soundtrack during familiarization was used. This was calculated in a manner similar to that described above, and these mean proportions are depicted in Table 1B.

Interobserver reliability for results of the intermodal learning test was calculated on the basis of 29 of the 100 subjects (at least five from each familiarization condition) and was .93. This is a Pearson product-moment correlation between the proportions of total looking time the infant spent fixating the sound-appropriate film as derived from observations of the primary and secondary observers separately.

Results of the intermodal learning test were examined to determine whether intermodal learning had occurred on the basis of the 2-min familiarization with single films and soundtracks. To address this question, results of the intermodal learning test for infants in each of the five familiarization conditions were compared. A one-way analysis of variance was performed on the proportions of total looking time infants spent fixating the sound-appropriate films (depicted in Table 1A) across the five conditions. Results revealed a significant main effect of familiarization condition on test performance, $F(4,95) = 4.93$, $p < .01$, indicating that the different familiarization conditions resulted in significantly different performance on the intermodal learning test.

Further analyses were conducted to determine which type of familiarization condition resulted in intermodal learning as compared with the performance of subjects in the control group. It is apparent from inspecting the means in Table 1A that subjects in the A/S condition displayed the strongest looking proportion (.62) to the sound-appropriate films during the intermodal learning test, while those of the other three experimental familiarization conditions displayed weaker looking proportions, close to that of the control group (.53). These mean looking proportions for the four experimental familiarization conditions were compared with that of the control group using Dunnett's t statistic. This method for comparing each treatment condition against a single control condition represents a somewhat conservative approach to handling the problem of multiple comparisons by adjusting the significance level to reflect that of all relevant comparisons. Results of this test indicated that the mean looking proportion of only the A/S group differed significantly from that of the control group, $t(95) = 3.01$, while comparisons of looking proportions between the A/N, I/S, and I/N conditions with the control group did not differ significantly, $t(95) = .49$, 1.02, .72, respectively. This summary decision has a significance level of .02. These results indicate that only subjects who were familiarized with appropriate and synchronous film and soundtrack pairs performed differently from control subjects on the intermodal learning test. They showed evidence of intermodal learning, whereas those who received familiarization with inappropriate or nonsynchronous film and soundtrack pairs showed no evidence of intermodal learning relative to performance of the control group subjects.

However, before this conclusion could be accepted unequivocally, another hypothesis was entertained. It could be argued that learning scores for subjects in the two nonsynchronous familiarization conditions (A/N, I/N) were artificially low because only synchronous soundtracks were presented during the test trials, and, in this respect, the test trials did not parallel the learning conditions. Therefore, an additional control group was run to test this possibility. Twenty subjects, matched for age with those of the other conditions, were randomly assigned to either the A/N or I/N control conditions. The procedure for familiarization and test was identical to that of their experimental counterparts, except that the soundtracks were played out of synchrony with the motions of both kinds of objects during the test trials. Results were comparable to those previously obtained with synchronous test trials. Subjects looked an average of .54 ($SD = .07$) in the A/N condition and .49 ($SD = .06$) in the I/N condition to the sound-specified film during the test. T tests indicated that these means were not significantly different from those of their experimental counterparts, nor were they different from that of the control group that received irrelevant familiarization ($p > .1$, all tests). Given no significant differences, the above conclusion, that only subjects in the A/S condition demonstrated intermodal learning, was accepted.

The data were also analyzed to determine whether subjects showed any side or film preferences during the intermodal learning test. Two two-way analyses of variance were performed on the proportions of total looking time infants spent fixating the right versus left film and the single versus compound film when each was sound specified. The analysis for side preference, with condition (A/S, A/N, I/S, I/N, and control) as a between-subjects factor and side (right vs. left) as a within-subjects factor, indicated no significant main effect of side, $F(1,95) = .15, p > .05$, or side × condition interaction, $F(4,95) = 1.44, p > .05$. The analysis for film preference, with condition as a between-subjects factor and film (single vs. compound object) as a within-subjects factor, indicated a nearly significant main effect of film, $F(1,95) = 3.60, p = .061$, and a nonsignificant film × condition interaction, $F(4,95) = 1.47, p > .05$. Across the five conditions, the film of the compound object was fixated approximately .58 of the time when it was sound specified, while the film of the single object was fixated approximately .47 of the time when it was sound specified. This preference for the film of the compound object was evident in three of the five conditions (A/N, I/S, and I/N); however, no film preference was evident for subjects in the A/S condition. They spent approximately .62 of their total fixation time viewing each film when it was sound specified ($SD = .27$ for compound object; $SD = .32$ for single object).

It should be noted that although the mean looking proportions to the sound-appropriate film (depicted in Table 1A) are an appropriate way of expressing results of the intermodal learning test for the purpose of comparing performance of the four experimental conditions against that of the control group, they are not appropriate for examining specific questions concerning the nature and direction of intermodal learning. For this purpose, the results are best conceptualized as depicted in Table 1B, in terms of the mean proportion of total looking time infants spent fixating the film that had previously been paired with the soundtrack during familiarization. This way of expressing the data allows one to examine the direction of the intermodal learning effects across the four experimental conditions. However, the test performance of control subjects cannot be expressed in this way since they did not receive familiarization with the test stimuli. Thus, all subsequent analyses of test performance are based on the proportions depicted in Table 1B.

Results of the learning test were analyzed to determine whether subjects showed any trends over trials in looking to the film that had previously been paired with the soundtrack during training. A two-way repeated-measures analysis of variance was performed, with condition (A/S, A/N, I/S, I/N) as a between-subjects factor and trials (1–8) as a within-subjects factor. Results indicated a significant main effect of condition, $F(3,76) = 10.10$, $p < .01$, and no significant main effect of trials, $F(7,532) = .55$, $p > .05$, or trial × condition interaction, $F(21,532) = .88$, $p > .05$. Thus infants' performance did not differ significantly across trials during the learning test. It neither declined nor improved significantly across trials.

On what basis did the intermodal learning of subjects in the A/S condition occur? To address this question, a two-way analysis of variance was performed to test for main effects of synchrony and sound appropriateness during familiarization on subsequent looking proportions of the intermodal test. Results indicated a significant main effect of synchrony, $F(1,76) = 13.05$, $p < .01$, and a significant main effect of sound appropriateness, $F(1,76) = 16.9$, $p < .01$. The interaction between synchrony and sound appropriateness was not significant, $F(1,76) = 1.52$, $p > .05$. These results indicate that infants showed intermodal learning on the basis of both synchrony and sound appropriateness.

To further assess the relative effects of synchrony and sound appropriateness on intermodal learning, pairwise comparisons were made using a Scheffé test for all pairs of means depicted in Table 1B. Results indicated that the learning-test performance of subjects who received appropriate and synchronous familiarization differed significantly from that of each of the other three groups ($p < .05$). No other pairs of means differed significantly. This suggests that both synchrony and sound appropriateness together had a significantly greater effect on intermodal learning than did either factor individually. This may be because asynchrony and sound inappropriateness have an interfering effect on intermodal learning.

Conclusions

In summary, when 3-month-old infants were given the opportunity to learn a relation between a single film and soundtrack under one of four familiarization conditions, only those subjects who received films paired with soundtracks that were synchronously related to the motions of the objects and appropriately related to the composition of the objects showed evidence of intermodal learning. Their mean looking proportion to the sound-specified film during the intermodal learning test was significantly greater than that of the control group, which had received irrelevant stimuli for familiarization. The test performance of subjects in the other three conditions was not significantly different from that of the control group.

Several conclusions can be drawn on the basis of these findings. First, 3-month-old infants were able to learn the relation between visible and acoustic stimulation by simply watching a single, naturalistic, audible and visible event. This intermodal learning took place on the basis of a single minute of exposure to each of two kinds of films and accompanying soundtracks. Because the test performance of subjects in the A/S condition was significantly better than that of the naive control subjects, it could not have been based on prior experience. Rather, it was the result of the 2-min familiarization period with these events.

Second, this intermodal learning was based on the infants' detection of amodal invariant relations. The infants abstracted two kinds of amodal relations that united the audible and visible stimulation from these filmed events during the 2-min familiarization period. This conclusion is supported by the finding that only subjects in the A/S condition showed evidence of intermodal learning, whereas those in the other three conditions did not. The A/S condition differed from the other three in only one way: it provided two levels of invariant audio-visual structure since the soundtracks were both synchronous with the motions of the objects and appropriate to their composition. Infants in the A/S condition must have detected the temporal macrostucture, or synchrony uniting the sights and sounds of impact, and the temporal microstructure available at each synchronous impact specifying the composition of the moving object: in this case whether the object was composed of a single, larger unit or an aggregate of smaller units. This interpretation was also confirmed by the finding of significant main effects of both synchrony and sound appropriateness on infants' performance during the intermodal learning test.

Third, infants apparently learned to perceive a relation between the co-occurring patterns of audible and visible stimulation when they detected invariant temporal structure uniting them, and they did not learn to perceive a relation between them when they detected any temporally discrepant audio-visual structure. Although infants in all four familiarization conditions were given the opportunity to learn a relation between a single film and soundtrack pair, only subjects who received synchronously and appropriately related film and soundtrack pairs showed evidence of learning. Subjects showed no evidence of learning to perceive a relation between the films and soundtracks when the soundtracks were either nonsynchronous or inappropriately related to the composition of the filmed object. The presence of incongruent information may have interfered with infants' learning on the basis of congruent bimodal temporal structure. These conclusions support and extend the findings of Bahrick (in press), where temporal asynchrony disrupted 6-month-olds' detection of bimodal temporal microstructure. The above findings suggest that by the age of 3 months, there are inherent constraints on the kinds of audio-visual relations infants are capable of learning. Just as early research on taste aversions in animals and humans has shown an inherent bias toward relating sickness with preceding tastes (Bernstein, 1978; Garcia & Koelling, 1966; Garcia, Rusiniak, & Brett, 1977), the present research, from a different theoretical perspective, suggests an inherent bias toward relating temporally congruent audible and visible stimulation.

In prior research, infants of 6 months or younger have demonstrated detection of temporal synchrony (Bahrick, 1983; Dodd, 1979; Lawson, 1980; Spelke, 1979, 1981; Spelke et al., 1983; Spelke & Cortelyou, 1980; Walker, 1982) and temporal microstructure (Bahrick, 1983, in press) uniting one of two simultaneously presented visual events with a soundtrack. The present research is consistent with the above findings of invariant detection and extends these findings to intermodal learning of natural object-sound relations by younger infants who previously showed no knowledge of the object-sound relations in question. Three-month-old infants who were unable to detect the relation between one of two simultaneously presented films and its soundtrack in a 2-min test period nevertheless were able to do so by first viewing the film alone, along with its appropriate and synchronous soundtrack. This perceptual learning must be similar to that which occurs in the natural environment, since this kind of bimodal stimulation is provided by most ordinary audible and visible events.

In contrast with the present findings, however, infants in several of the above studies were able to learn to relate objects with sounds that those objects could not naturally produce and that shared no temporal microstructure with their motions. That is, infants were able to disregard inconsistent temporal microstructure and learn to pair an object and sound on the basis of other information such as common tempo of action (Spelke, 1979), temporal synchrony (Lawson, 1980; Spelke, 1978, 1979, 1981), or even simple audio-visual co-occurrence (Lawson, 1980). In Spelke's research, 4-month-olds learned to relate the sight of a bouncing stuffed animal with the discrete percussive sound of a tennis shoe hitting a wooden box, or an oil-drum lid. Lawson (1980) even found that 6-month-olds learned to relate the sound of a piano note or a squeaky toy with a moving form that changed its trajectory periodically. This contrasts with the present findings and with results of Bahrick (1983) where infants showed no evidence of learning to relate a moving object and sound that were inappropriate, even when they were temporally synchronous, but were able to relate the natural and appropriate sound with the synchronously moving object. In these studies, however, the inappropriate sounds presented were perhaps more extremely mismatched with the motions of objects than those of Spelke or Lawson. They specified an object of a different substance (rigid vs. elastic) or an object of a different composition (single vs. many objects). It may be that the above studies, taken together, roughly delineate the boundaries of infants' capabilities for relating certain types of sounds with certain types of visual transformations: by 4½ months, the impact or trajectory change of a variety of single, fairly rigid objects can be related to a number of discrete, percussive impact sounds and even with discrete musical tones or squeaks, but they cannot be related to sounds that specify an object of an elastic substance, or a group of smaller objects together, or to continuous sounds. On the other hand, the stimulus-presentation format used by Bahrick (1983) and in the present research may have enhanced infants' perception of audio-visual temporal microstructure, since it was designed to examine just that. Unlike the other studies, both objects that naturally produced each type of sound were visible simultaneously or in successive trials, and the nature of their impacts could be compared and contrasted and related to the nature of the ongoing sound. In the other studies, the objects that naturally produced these sounds were not presented, and these invariants were totally absent. Further research must establish the boundaries of the infant's capability for relating specific types of sounds with specific types of motions and the nature of this developmental change.

The present research also sheds some light on the origins of intermodal learning in infancy. It delimits to some extent the nature and basis of infants' learning about the visual properties of acoustically specified events. The current results are consistent with Gibson's (1969) invariant-detection view of perceptual development and are inconsistent with at least one common interpretation of the association-integration view. This view is that infants associate sights and sounds on the basis of temporal contiguity such that when any audible and visible stimulation occur together within some brief, unspecified period of time, the two kinds of stimulation would be associated and integrated. However, in the present study, by 3 months of age, intermodal learning occurred on the basis of invariant detection and did not occur on the basis of simple co-occurrence between two patterns of stimulation. Infants in all four familiarization conditions had an equal opportunity for associating a co-occurring visual event and soundtrack, while only those in the appropriate and synchronous familiarization condition showed evidence of learning a relation between them. In all four familiarization conditions, each time an impact sound was heard, only one object was visible at that moment. In some conditions (A/N and I/N), at the moment of sound the object moved in a manner that was unrelated to the timing of the sound, and in other conditions (A/S and A/N) a trajectory change and visual impact were visible at the moment of sound. Even the synchronous co-occurrence of a sound and visible impact was not sufficient to promote learning. This version of the association-integration view thus does not adequately account for the differential learning across the four familiarization conditions. Thus, by 3 months of age, the origins of intermodal learning about the present kinds of naturalistic events are rooted in the detection of two types of amodal temporal invariants.

By using an intermodal learning method similar to the present one, future research may reveal more about the nature and basis of the development of intermodal perception. It may be used to explore which kinds of invariant relations are detected first and how this leads to differentiation of other invariants and to learning about more specific object-sound relations. The inclusion of a control group allows the researcher to examine directly the learning process in infants who are too inexperienced to show evidence of intermodal perception in other common methods. Further, it permits direct control and manipulation of the learning context prior to assessing intermodal learning. These factors together promise to bring us closer to an understanding of how young perceivers develop the intermodal knowledge characteristic of adult perceivers.

References

Allen, T. W., Walker, K., Symonds, L., & Marcell, M. (1977). Intrasensory and intersensory perception of temporal sequences during infancy. *Developmental Psychology*, **13**, 225–229.

Bahrick, L. E. (1983). Infants' perception of substance and temporal synchrony in multimodal events. *Infant Behavior and Development*, **6**, 429–451.

Bahrick, L. E. (in press). Infants' intermodal perception of two levels of temporal structure in natural events. *Infant Behavior and Development*.

Bernstein, I. L. (1978). Learned taste aversions in children receiving chemotherapy. *Science*, **200**, 1302.

Birch, H. G., & Lefford, A. (1963). Intersensory development in children. *Monographs of the Society for Research in Child Development*, **28** (5, Serial No. 89).

Birch, H. G., & Lefford, A. (1967). Visual differentiation, intersensory integration, and voluntary motor control. *Monographs of the Society for Research in Child Development*, **32**(1–2, Serial No. 110).

Blank, M., & Bridger, W. H. (1964). Crossmodal transfer in nursery school children. *Journal of Comparative and Physiological Psychology*, **58**, 277–282.

Bower, T. G. R., Broughton, J. M., & Moore, M. K. (1970a). The coordination of visual and tactual input in infants. *Perception and Psychophysics*, **8**, 51–53.

Bower, T. G. R., Broughton, J. M., & Moore, M. K. (1970b). Demonstration of intention in the reaching behavior of neonate humans. *Nature*, **228**, 679–680.

Bryant, P. (1974). *Perception and understanding in young children: An experimental approach.* London: Methuen.

Bushnell, E. W. (1981). The ontogeny of intermodal relations: Vision and touch in infancy. In H. L.

Pick, Jr., & R. D. Walk (Eds.), *Intersensory perception and sensory integration* (pp. 5–36). New York: Plenum.

Bushnell, E. W. (1986). The basis of infant visual-tactual functioning—amodal dimensions or multimodal compounds? In L. P. Lipsitt & C. Rovee-Collier (Eds.), *Advances in infancy research* (Vol. 4, pp. 182–194). Norwood, NJ: Ablex.

Dodd, B. (1979). Lip reading in infants: Attention to speech presented in-out-of-synchrony. *Cognitive Psychology*, **11**, 478–484.

Garcia, J., & Koelling, R. A. (1966). The relation of cue to consequence in avoidance learning. *Psychonomic Science*, **4**, 123–124.

Garcia, J., Rusiniak, K. W., & Brett, L. P. (1977). Conditioning food-illness aversions in wild animals: Caveant canonici. In H. Davis & H. M. B. Hurwitz (Eds.), *Operant Pavlovian interactions* (pp. 273–311). Hillsdale, NJ: Erlbaum.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.

Gibson, E. J., & Walker, A. S. (1984). Development of knowledge of visual-tactual affordances of substances. *Child Development*, **55**, 453–460.

Gottfried, A. W., Rose, S. A., & Bridger, W. H. (1977). Cross-modal transfer in human infants. *Child Development*, **48**, 118–123.

Lasky, R. E. (1977). The effect of visual feedback of the hand on the reaching and retrieval behavior of young infants. *Child Development*, **48**, 112–117.

Lawson, K. R. (1980). Spatial and temporal contiguity and auditory-visual integration in infants. *Developmental Psychology*, **16**, 185–192.

Lyons-Ruth, K. (1977). Bimodal perception in infancy: Response to auditory-visual incongruity. *Child Development*, **48**, 820–827.

Meltzoff, A. N., & Borton, R. W. (1979). Intermodal matching by human neonates. *Nature*, **282**, 403–404.

Mendelson, M. J., & Ferland, M. B. (1982). Auditory-visual transfer in four-month-old infants. *Child Development*, **53**, 1022–1027.

Ruff, H. A., & Kohler, C. J. (1978). Tactual-visual transfer in six-month-old infants. *Infant Behavior and Development*, **1**, 259–264.

Spelke, E. S. (1976). Infants' intermodal perception of events. *Cognitive Psychology*, **8**, 553–560.

Spelke, E. S. (1978). *Intermodal exploration by four-month-old infants: Perception and knowledge of auditory-visual events*. Unpublished doctoral dissertation, Cornell University.

Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology*, **15**, 626–636.

Spelke, E. S. (1981). The infant's acquisition of knowledge of bimodally specified events. *Journal of Experimental Child Psychology*, **31**, 279–299.

Spelke, E. S., Born, W. S., & Chu, F. (1983). Perception of moving, sounding objects by four-month-old infants. *Perception*, **12**, 719–732.

Spelke, E. S., & Cortelyou, A. (1980). Perceptual aspects of social knowing: Looking and listening in infancy. In M. E. Lamb & L. R. Sherrod (Eds.), *Infant social cognition* (pp. 61–84). Hillsdale, NJ: Erlbaum.

Walker, A. S. (1982). Intermodal perception of expressive behaviors by human infants. *Journal of Experimental Child Psychology*, **33**, 514–535.