JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 53, 180–199 (1992)

Infants' Perceptual Differentiation of Amodal and Modality-Specific Audio–Visual Relations

LORRAINE E. BAHRICK

Florida International University

Ninety-six infants of 3½ months were tested in an infant-control habituation procedure to determine whether they could detect three types of audio–visual relations in the same events. The events portrayed two amodal invariant relations, temporal synchrony and temporal microstructure specifying the composition of the objects, and one modality-specific relation, that between the pitch of the sound and the color/shape of the objects. Subjects were habituated to two events accompanied by their natural, synchronous, and appropriate sounds and then received test trials in which the relation between the visual and the acoustic information was changed. Consistent with Gibson's increasing specificity hypothesis, it was expected that infants would differentiate amodal invariant relations prior to detecting arbitrary, modality-specific relations. Results were consistent with this prediction, demonstrating significant visual recovery to a change in temporal synchrony and temporal microstructure, but not to a change in the pitch–color/shape relations. Two subsequent discrimination studies demonstrated that infants' failure to detect the changes in pitch–color/shape relations could not be attributed to an inability to discriminate the pitch or the color/shape changes used in Experiment 1. Infants showed robust discrimination of the contrasts used.
© 1992 Academic Press, Inc.

Research in the area of intermodal perception has mushroomed during the past 15 years and revealed many interesting intermodal capabilities in infancy. Infants by 4 months can detect amodal invariant relations uniting the audible and visible stimulation in unitary events. They detect temporal synchrony (Bahrick, 1983; Spelke, 1979) and tempo of action (Spelke, 1979) common to visually and acoustically specified impacts of

This research was supported by National Institute of Child Health and Human Development Grants 5-R23-HD18766 and 1-RO1-HD25669 to the author. A portion of these data were presented at the Fifth International Conference on Event Perception and Action in Oxford, Ohio, July, 1989. Special appreciation is extended to Isabel Alfonsin, Alicia Cuesfa, Maria Hernandez, Laura Mayorga, Jeannie McKinney, Gloria Peruyera, and Jeffrey Pickens for their assistance in the collection and analysis of data. Requests for reprints should be addressed to Lorraine Bahrick, Department of Psychology, Florida International University, Miami, FL, 33199.

0022-0965/92 $3.00
Copyright © 1992 by Academic Press, Inc.
All rights of reproduction in any form reserved.

an object against a surface. They also detect audio–visual temporal information specifying the substance and the composition of an object striking a surface (Bahrick, 1983, 1987, 1988). By 5 to 7 months they abstract the common affective expression uniting the voice and face of a speaking person (Walker, 1982; Walker-Andrews 1986), and at 4 months the relation between the shape of the lips and the corresponding vowel sound in speech (Kuhl & Meltzoff, 1984). Few studies, however, have attempted to assess the developmental progression of infants' detection of different intermodal relations. How does detection of intermodal relations of one type lead to detection of other relations and eventually to the veridical knowledge about intermodal events possessed by adult perceivers? The present study is a first step in this direction. It assesses the ability of 3½-month-old infants to abstract three audio–visual relations available in a single event.

A particular audible and visible event makes many audio–visual relations available. Some relations are global and others more specific; some are amodal and invariant (Gibson, 1969), while others are arbitrary and not invariant across sense modalities. For example, when an object repeatedly falls against a surface, amodal relations such as temporal synchrony unite the sights and sounds of the impacts, and temporal microstructure specifies the composition or substance of the object (whether it is rigid or elastic, a single or compound object, see Bahrick, 1983, 1987). Arbitrary relations such as that between the color or shape of the object and the pitch of its impact sound are also available. There is little, if any, predictable relation between the color/shape of an object and the pitch of its impact sound. Which multimodal relations can be appropriately generalized to other similar events? Are these relations detected first, while those that would generalize inappropriately detected later? These fundamental questions have received no empirical testing to date.

According to Gibson (1969) infants are predisposed to detect amodal invariant relations and abstraction of these relations guides exploration of multimodal events and the development of intermodal knowledge. Perceptual learning is thought to proceed in order of increasing specificity. By detecting amodal invariant relations infants are able to determine which visible object is the source of the sound they are hearing and can use this information to guide further exploration of the unitary event. Knowledge about amodal invariant relations may be appropriately generalized to other events. For example, by abstracting amodal temporal relations specifying the substance of an object, infants may learn that hard objects make rigid *perspective transformations* and produce sounds that are short and abrupt, while those that are soft produce deforming motions and make sounds that are more prolonged and have a more gradual onset time. This learning may facilitate exploration and learning about other rigid and elastic objects. However, if infants detect arbitrary, modality-specific relations such as the relation between the pitch of a sound and the color or shape of an object, inappropriate generalization to new events may result.

What distinguishes amodal from arbitrary relations? Amodal relations are those that provide completely redundant information to two or more sense modalities (Gibson, 1969, p. 219). For example, the rhythm or tempo of an audible and visible event can be abstracted from either the visual or the acoustic stimulation. Detection of amodal relations is assumed to require no prior learning (Gibson, 1969). In contrast, arbitrary relations are defined here as those that provide no overlapping or invariant information across sense modalities, and thus detection of these relations requires learning.

Walker-Andrews, Bahrick, Raglioni, and Diaz (1991) distinguished between two types of arbitrary intermodal relations; those that are arbitrary and natural, and those that are arbitrary and artificial. Arbitrary, natural relations are cross-modal relations that predictably occur together in the environment and have natural, built-in constraints. They include the relation between the sight of a person's face and the particular sound of their voice, or the sight of a person and their characteristic odor. There are inherent constraints with respect to the types of sounds people can produce. Research has shown that by the age of 3½ months, infants have already learned the association between the sight of their mother's face and the sound of her voice (Spelke & Owsley, 1979).

In contrast, arbitrary, artificial relations are those that do not predictably occur together in nature and are relatively unconstrained. For example, the relation between the color or shape of an object and its temperature, taste, or the type of sound it produces. A yellow, square object could as easily emit a hum or a squeak, produce an impact sound of a high or a low pitch, or be sweet, sour, warm, or cold. These relations can only be learned through experience with the particular objects and they may even change from one context to another. The temperature of an object will change depending on the temperature of the environment; the pitch of an impact sound will change depending on the surface against which the object is struck. Given such limited generalizability, detection of these relations seems less important for perceptual development than detection of amodal or arbitrary, natural relations. Only a few studies have explored infants' detection of these artificial, arbitrary relations. Reardon and Bushnell (1988) found that 7-month-olds were able to learn the relationship between the color of a container and the taste of its food. On the other hand, Bushnell (1986) found that infants were unable to learn to pair the color of an object with its temperature. Learning about arbitrary relations such as these may occur through association on the basis of cooccurrence.

Alternatively, intermodal knowledge about arbitrary relations may also be acquired by first detecting amodal invariant relations. Some studies have demonstrated that detection of amodal relations leads to intermodal

knowledge about arbitrary, artificial, audio–visual relations. Spelke (1979, 1981) found that 4-month-olds were able to learn the relation between the appearance of a stuffed toy (kangaroo versus donkey) and its sound (thud or gong) when the motions and sounds were united by temporal synchrony or common tempo of action, but not when the object's motions were unrelated to the sounds. During the intermodal preference phase infants detected the amodal synchrony and tempo relations uniting each event with its characteristic sound even though the other event was simultaneously visible. Learning of arbitrary relations was demonstrated when infants were subsequently able to search for the appropriate object upon hearing its distinctive sound. Similarly, Lyons-Ruth (1977) found that infants learned to pair a sound with a distinctive toy when the sound was synchronized with the motions of the toy. Infants subsequently looked more to the familiar toy in the presence of the sound than to a novel toy. Lawson (1980) also found that infants learned about the relationship between the appearance of an object and a sound when the sound and object motions were temporally synchronized. In these studies, learning about the arbitrary relation between a sound and the appearance of an object was based on detection of amodal relations uniting the sounds with the object's motions. Detection of amodal relations may, in this manner, lead to intermodal knowledge about arbitrary, modality-specific relations. In this respect, it seems adaptive for infants to developmentally detect amodal relations prior to abstracting arbitrary, artificial relations. To what extent these arbitrary relations could be learned without first detecting amodal relations specifying the unity of the audible and visible stimulation is not known, although Lawson (1980) found some evidence of association at 6 months on the basis of spatial cooccurrence alone. Some arbitrary pairings of sights and sounds, however, are not united by amodal relations and must therefore be learned on the basis of association. Nevertheless, the adult perceiver has no trouble detecting and learning about these relations. At what point developmentally does this ability emerge, and is it preceded by the ability to detect amodal relations?

The present study addresses this question by presenting infants with single events that made audio–visual relations of three different levels of specificity available. They include two amodal relations and one arbitrary, artificial relation. Single and compound objects were shown striking a surface in an erratic pattern. (1) Temporal synchrony between the sights and sounds of impact could be abstracted. This is a global, amodal audio–visual relation. (2) Temporal microstructure specifying the composition of the object (whether it was composed of a single element or a group of smaller elements) could be abstracted. This is an amodal audio–visual relation of greater specificity than synchrony. For each synchronous impact, the nature of the sound and visual trajectory change specifies the composition of the object. A single, short sound and trajectory change specify a single, unitary element, while a more prolonged progression of sounds and a more prolonged visual trajectory change specify a cluster of elements striking a surface. (3) Modality-specific, arbitrary relations between the pitch of the sound and the color and shape of the object could be abstracted. These relations are considered to be of the greatest specificity. For each synchronous impact and for an object of a given composition, one can detect a particular pitch–color/shape pairing.

Consistent with an increasing specificity view, it was hypothesized that abstraction of amodal invariant relations would developmentally precede abstraction of arbitrary, modality-specific relations. Based on prior research (Bahrick, 1988), it was expected that infants at 3½ months would detect amodal synchrony and possibly composition relations. However, they may show no evidence of detecting the arbitrary pitch–color/shape relations at this age, even if the changes in color/shape and pitch were discriminable to them.

Infants' detection of synchrony relations and information specifying object composition was previously assessed in our lab using a version of Spelke's (1976) intermodal preference method. Bahrick (1987) found that it was not until 6 months of age that infants could determine which of two films belonged with a soundtrack they were hearing on the basis of audio–visual relations specifying the composition of the objects. There was some evidence for detection of synchrony by 4½ months, but not earlier. Further, Bahrick (1988) found that although 3-month-olds could not detect the relation between a soundtrack and one of two simultaneously presented films on the basis of either temporal synchrony or composition information, they were able to do so when training was first conducted with a single film and soundtrack that provided both types of invariant relations.

In the present study, an infant-control habituation procedure was used to assess detection of audio–visual relations in 3½-month-old infants. Since infants of this age were able to abstract both amodal synchrony and composition information during the training phase of the Bahrick (1988) study, where they received single film and soundtrack pairs, but were unable to do so when two films were presented along with a soundtrack that matched only one of them (Bahrick, 1987; and a prior pilot study, see General Discussion), the intermodal preference method was not used in the present research. It was hoped that the habituation procedure would provide a more sensitive measure. Like the training phase of the Bahrick (1988) study, it presents infants with only a single audible and visible event. Detection of a change in the audio–visual relations can then be assessed through the test trials. Thus, although the habituation method may be a more sensitive index of intermodal functioning, it cannot reveal whether infants are then able to use these intermodal relations to guide exploration when more than one event is visible. Evidence of intermodal

capabilities may thus be found at a younger age by using the habituation method since discrimination abilities may developmentally precede matching abilities in this domain.

EXPERIMENT 1

Method

Subjects. Ninety-six infants (44 males and 50 females) of 3½ months, whose mean age was 111.8 days (SD = 4.5 days), composed the final sample. Thirty-four additional subjects were unable to complete the testing procedure, 23 (whose mean age was 109.8 days, SD 5.5) because of excessive fussiness or crying, 7 because of equipment failure or experimenter error, and 4 who fell asleep. Subjects were recruited through the local birth records, and all were healthy with no complications upon delivery. The sample for this and the two subsequent studies was primarily middle class and parents were required to have at least 12 years of education.

Stimulus materials. Color, video films were made of six pairs of naturalistic audio–visual events. The events were chosen so as to portray two types of object composition. One member of each pair depicted a single, large, object (single object), and the other a group of smaller objects (compound object) of the same color and shape as the large one (see Fig. 1). All objects were filmed striking a surface in an erratic pattern and their natural impact sounds were heard. Each single object produced a single, abrupt, impact sound, while each compound object produced a slightly more prolonged sound with a more gradual rise time. Thus, the sounds and visible motions of the object were united by both a common temporal synchrony and temporal microstructure specifying object composition. The six event pairs composed three different event sets. Each set was constructed from a different material, metal, plastic, or wood, and depicted a different type of impact motion. Further, the two event pairs within a set differed only in color and shape from one another. These sets consisted of the following:

Set A: Metal objects. Event pair A_1: A single, large, yellow, metal washer suspended from a string and a group of small, yellow washers suspended from a string were raised and abruptly dropped against a wooden surface. Event pair A_2: A single, large, orange, metal nut suspended from a string and a group of small, orange, metal nuts suspended from a string were raised and dropped against a wooden surface.

Set B: Plastic objects. Event pair B_1: a single large, yellow, pear, held from behind by a small stick with an unseen hand and a group of small, yellow pears, also held from behind, were abruptly struck against two surfaces of different heights in an erratic back and forth motion. Event pair B_2: A single, large, red tomato, held from behind and a group of small red tomatoes were also struck against two surfaces as before.

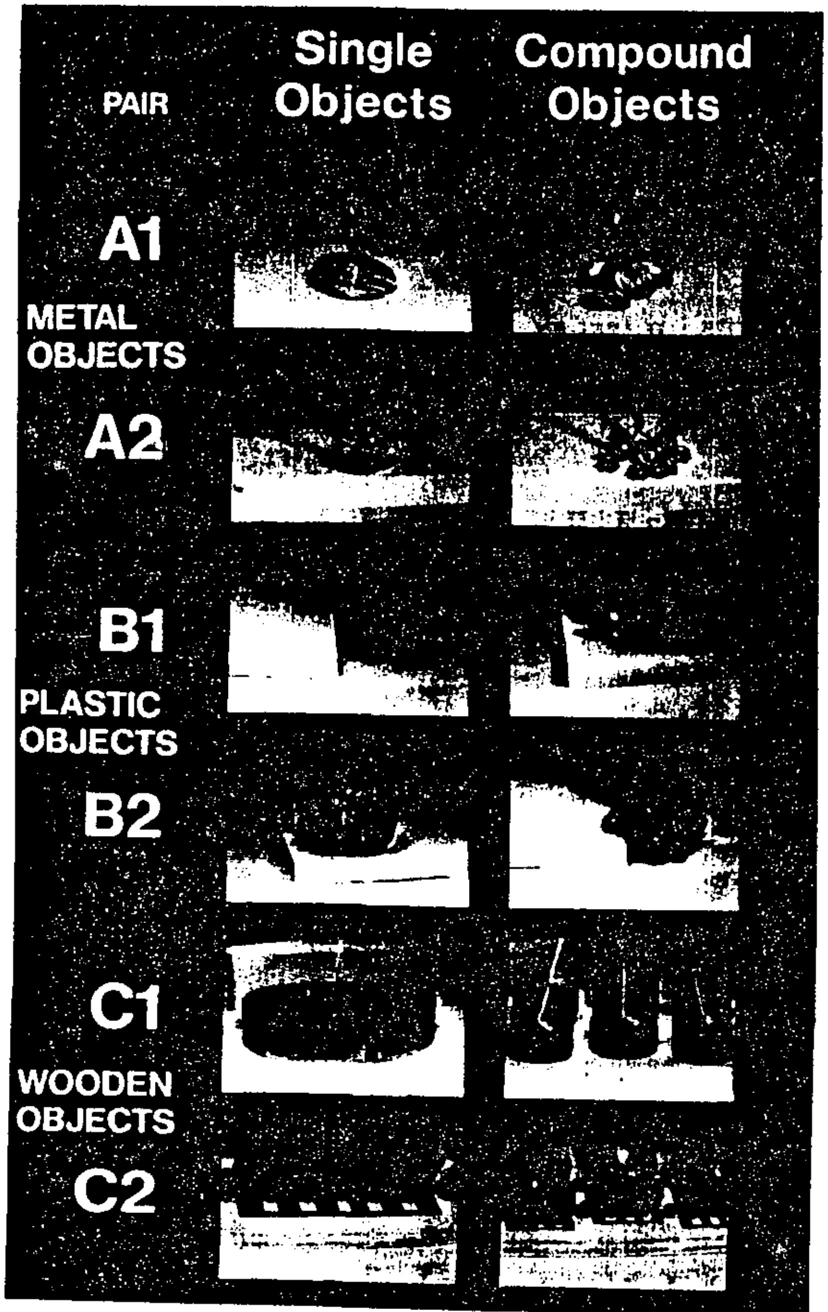

FIG. 1. Photograph of the stimulus events.

Set C: Wooden objects. Event pair C_1: A single, large, semicircular board with pink and purple stripes was hinged to a horizontal wooden surface and opened and slammed shut with a string from above. The compound object depicted a group of small wooden objects of the same color and shape as above, each hinged to a wooden surface and opened and slammed shut together (with a common string from above). Event pair C_2: A single, large, rectangular board with yellow and blue stripes and a group of smaller, rectangular, wooden objects with yellow and blue stripes were also hinged to a wooden surface and opened and slammed shut with a string from above.

All objects from each event pair struck the surface in an erratic temporal pattern and there were approximately 40 impacts per minute across the 5-min video films. Each pair of events portrayed two levels of nested amodal invariant temporal structure: temporal synchrony across impacts and temporal microstructure specifying object composition within each synchronous impact. A third audio–visual relation was also available. This was the arbitrary, modality-specific relation between the pitch of the sound and the color/shape of the object and was considered to be the relation of greatest specificity. It is referred to as the specific object–sound relation. This was created by electronically altering the natural impact sound of each event such that in one version the sound was approximately four whole notes higher and in the other four whole notes lower than the natural sound. The two sounds thus differed by at least an octave. Thus, each event could be characterized as having three audio–visual relations.

In addition to the primary stimulus displays, a videotaped display was made to serve as a control stimulus. The event depicted a green and white plastic toy turtle, whose front legs spun around, creating a whirring sound.

Apparatus. Infants were seated in an infant seat facing a 19-in. (Panasonic BT-S1900N) video monitor, approximately 55 cm away. A set of colored Christmas tree lights and a mechanical toy doy were positioned above and just to the right of the video display. The video monitor was surrounded by black posterboard. Two apertures, one to the upper right and the other to the upper left side of the video monitor allowed observers to monitor the subjects' visual fixations.

The stimulus events were videotaped with a Panasonic WV 3170 color video camera and a Sony EMC-150T remote microphone. They were then edited and presented with a Panasonic VHS NV-A500 edit controller connected to two Panasonic video decks (NV-8500 and AG-6300). The use of two video decks allowed us to switch from one display to another on the same monitor without extra time or noise resulting from changing cassettes. All soundtracks were presented through a speaker positioned just below the video monitor at approximately 65–70 db.

For those conditions where the soundtracks were modified, an auditory signal processor (Yamaha SPX 90) raised or lowered the pitch of the impact sound by approximately 3.5–4 whole steps. Thus, the high-pitched sounds and low-pitched sounds for each event were approximately one octave apart. Twenty adult observers judged both the high and the low sounds to be plausibly produced by the visual events.

For conditions where asynchronous sounds were presented, the natural soundtrack was dubbed onto the video film so that it was out of phase with the object's motions and unsystematically related to them. Inappropriate but synchronous soundtracks for each event were also created. By using a soundproof box, one visual event was filmed while the sounds of a different synchronous event were simultaneously recorded.

A trained observer who was unaware of the subject's condition (experimental versus control) monitored visual fixations by depressing a button while the infant fixated the video image. The button box was connected to an Apple IIe personal computer which was programmed to record visual fixations on-line, signal when the infant had looked away for a given amount of time, and signal when the infant had reached a preset habituation criterion. The signal was transmitted through a small speaker and headphone to the experimenter who operated the video equipment from a table behind the infant seat. A permanent record of the infant's visual fixation pattern was printed during the experiment using a Qume Letterpro 20-S printer. Only the observations of the primary observer controlled the stimulus presentation, while the observations of a secondary observer were simply recorded by the computer and used for later calculation of interobserver reliability.

Procedure. Subjects were tested using an infant-control habituation procedure (Horowitz, Paden, Bhana, & Self, 1972) to determine whether they could detect a change in synchrony, composition, or specific object–sound relations. Ninety-six infants were randomly assigned to one of three experimental conditions, synchrony, composition, or specific object–sound, or one of three corresponding control conditions. For synchrony and composition, 18 subjects were assigned to each experimental condition and 12 to each control condition. For specific object–sound, 24 subjects were assigned to the experimental and 12 to the control condition. During each experimental condition, subjects were habituated to two events, in an alternating sequence, with their synchronous and appropriate sounds. Then for the test trials they received the same two events, however, the auditory and visual information were presented in a novel relation to one another. Thus, during the test trials for all experimental conditions, both the films and their soundtracks were familiar while only the relation between them was novel.

The three experimental conditions were conducted as follows: (1) Synchrony: Subjects were habituated to two events with synchronous sounds and tested with nonsynchronous sounds. The nonsynchronous sounds were created by playing the natural soundtracks out of phase with the videos, such that the sounds were unsystematically related to the visual impacts of the objects. The two events consisted of a single and a compound object from the same event set, but of a different color and shape. For example, one subject might receive the large yellow pear and the small red tomato events accompanied by synchronous sounds for habituation and the same nonsynchronous sounds for test. Thus, only synchrony was changed from habituation to test. (2) Composition: Subjects were habituated to displays of a single and a compound object from the same event set and of the same color and shape (e.g., large yellow pear vs small yellow pears). During habituation the sounds and films were synchronous and appropriate and during the test they were synchronous but mismatched. Thus, the motions of the single object were synchronized with the sounds of the compound object, while the motions of the compound object were synchronized with the sounds of the single object. Only the relation between the type of sound and the type of object changed from habituation to test. (3) Specific object–sound: Subjects were habituated to displays of either two single or two compound objects from the same event set, differing only in color and shape. One object produced a natural sounding high-pitched impact sound, while the other an impact sound of a lower pitch. Test trials consisted of a change in the pairing of pitch with color/shape. For example, a subject might receive the large yellow pear with the high sound and the large red tomato with the low sound for habituation. For test trials the pear would be played with the low sound and the tomato with the high sound. Since originally these objects had essentially the same sound when they struck the surface, the assignment of a high or low sound was arbitrary and counterbalanced across subjects. From habituation to test, only the pairing of pitch with the object's color and shape changed.

Subjects in the control conditions received an habituation sequence identical to that of their experimental counterparts, however, there was no change from habituation to test.

Stimulus set and event pair were counterbalanced within each experimental and control condition. One-third of the subjects within each condition received an event pair from the set of metal objects, one-third from the wooden objects, and one-third from the plastic objects. Within each of these cells half the subjects received one of the two possible event pairs, and the other half the other possible event pair. For each subject, it was randomly determined which of the two events would be presented first. Presentation order then followed an alternating sequence. In addition, for subjects in the SOS condition, the assignment of high- and low-pitched sounds to objects was also counterbalanced.

Prior to the habituation sequence and just after the test trials, a control display was presented to serve as a warm-up period and as a check for fatigue. The observer's recording of this trial and all habituation and test trials was initiated as soon as the infant fixated the visual display and was terminated once the infant had looked away for 1.5 s. In addition, a ceiling of 60 s was set as the maximum trial length. The intertrial intervals were approximately 6 s long and the colored lights were illuminated during this time. The mechanical toy doy was only infrequently activated, always at some point prior to the posthabituation trials, to encourage subjects to attend to task.

The habituation sequence in general consisted of four mandatory trials and was then terminated as soon as the infant reached the habituation criterion. The criterion was defined as a fixation decrement of 50% or greater on two consecutive trials, relative to the infant's initial fixation level (the average number of seconds of fixation during the first two habituation trials). After the two criterion trials, two (no-change) posthabituation trials were presented. These trials served to establish a more conservative habituation criterion, reducing chance habituation and taking into account spontaneous regression effects (see Bertenthal, Haith, & Campos, 1983, for a discussion of these effects). Following the habituation sequence, subjects received two test trials, one of each event, to assess visual recovery and then a single trial during which the control display was again presented. The two test trials were presented in the same order as the first two events seen for habituation.

The subject's data were examined to determine whether two criteria had been met. To ensure that subjects had in fact habituated to the displays, data from subjects whose mean posthabituation fixation level exceeded that of their mean initial fixation level (baseline) were excluded from the study ($N = 12$). Second, to ensure that infants were not overly fatigued, and thus unable to show visual recovery, their fixation duration in the final control trial was compared with that of the initial control trial. If the final level was less than 20% of the initial level, the subject's data was also omitted ($N = 11$). The remaining subjects in the sample showed substantial visual recovery on the final control trial (median = 99% of initial fixation level).

A secondary observer monitored visual fixation for 18% of the sample for assessing interobserver reliability. For each infant, total fixation time on each trial was calculated independently on the basis of observations made by the primary and secondary observers. A Pearson product-moment correlation between these observations served as the measure of reliability and was 99.

Results

Primary analyses. Visual recovery to the two test displays was calculated by subtracting the average fixation time across the two posthabituation trials from that of the two test trials for each infant. This difference score thus reflects the change in fixation time to the two test displays relative

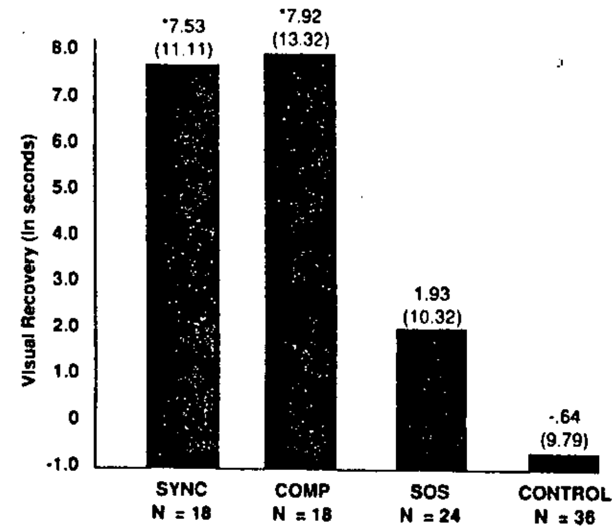

FIG. 2. Mean visual recovery and standard deviations for synchrony (SYNC), composition (COMP), specific object–sound (SOS), and control conditions. (*p = or < .01 with respect to control group.)

to the subject's fixation time to the two habituation displays just after the habituation criterion was met. Recovery scores were averaged across subjects within each condition to obtain a mean visual recovery score. Further, the visual recovery scores of subjects in the three control conditions (synchrony, composition, and specific object–sound) were collapsed into a single control condition ($N = 36$) for comparison with the experimental conditions, after an analysis of variance revealed no significant main effect of condition on visual recovery scores ($F(2, 33) = 1.09$, $p > .1$), nor on any of five secondary measures (see secondary analyses section for details). Figure 2 displays the mean visual recovery scores for each of the four conditions. An analysis of variance was conducted on the mean recovery scores across the four conditions (synchrony, composition, specific object–sound, and collapsed control) and indicated a significant main effect of condition ($F(3, 92) = 3.67$, $p < .02$). Planned t tests comparing the mean visual recovery of each of the experimental conditions against that of the collapsed control condition revealed a significant visual recovery to the change in synchrony relations ($t(52) = 2.76$, $p < .01$) and to the change in composition relations ($t(52) = 2.68$, $p = .01$), but not to the change in specific object–sound relations ($t(58) = .97$, $p > .1$) with respect to the recovery observed in the control condition. Thus, 3½-month-olds detected the change in synchrony and composition relations, but not the change in the pairing of pitch with color/shape.

Further analyses were conducted to determine whether visual recovery differed as a function of stimulus set. A two-way analysis of variance with condition (synchrony, composition, specific object–sound, and control) and stimulus set (wood, metal, plastic) as between-subjects factors revealed no main effect of stimulus set ($F(2, 84) = .92, p > .1$) or interaction of stimulus set with condition ($F(6, 84) = .75, p > .1$).

Secondary analyses. The performance of subjects during the habituation phase was evaluated with respect to five measures (1) baseline, defined as the average length of fixation on the first two habituation trials, (2) mean number of trials required to reach the habituation criterion, (3) mean number of seconds to habituation, summed across trials, (4) average length of fixation on the two criterion trials, and (5) average length of fixation on the two (no-change) posthabituation trials.

Analyses of variance were conducted separately for each of these measures to determine whether the control conditions for synchrony, composition, and specific object–sound differed from one another as a function of any of these variables. Results indicated no significant main effect of condition for any of the measures, providing further justification for collapsing across these conditions ($p > .1$, all tests). Further, analyses of variance assessing the main effect of condition (synchrony, composition, specific object–sound, and collapsed control) were performed separately on each of the five measures. These analyses were conducted to determine whether initial interest in the stimuli, final interest level, or pattern of habituation differed across the four conditions prior to introducing the test trials. Results of all five analyses indicated no significant main effects, $p > .1$, all tests.

Discussion

Results indicated that although subjects showed no a priori differences across groups in their attention to the habituation displays, they showed differential visual recovery to the test displays as a function of their condition. Subjects who received a change in synchrony relations and those who received a change in composition relations demonstrated significant visual recovery, whereas those who received a change in the pairing of pitch with color/shape showed no significant visual recovery relative to the performance of control subjects. At 3½ months, infants detected a change in two kinds of amodal audio-visual relations, but not a change in an arbitrary, modality-specific relation. This pattern of results is consistent with the increasing specificity hypothesis suggesting that abstraction of amodal audio–visual relations developmentally precedes abstraction of modality-specific relations. However, an alternative hypothesis could also be entertained. Perhaps infants failed to detect a change in the pitch–color/shape relations because they were unable to discriminate between the high and low pitches of each sound and/or between the two

objects on the basis of shape and color differences. To detect the change in the pairing of pitch with color/shape, infants must be able to discriminate the visual and acoustic changes. Experiments 2 and 3 were thus conducted to determine whether 3½-month-olds could discriminate changes in the auditory and visual components of the events used in this study.

EXPERIMENT 2: AUDITORY DISCRIMINATION

This study assessed the ability of 3½-month-olds to discriminate the auditory changes presented in Experiment 1. Each of the 12 events used in Experiment 1 was presented to one experimental and one control subject for habituation. Experimental subjects received the familiar visual event with a change in pitch for test trials while controls received no pitch change.

Method

Subjects. Twenty-four subjects, 13 males and 11 females, whose mean age was 112.5 days ($SD = 7.2$ days) participated in the study. The data from 6 additional subjects were rejected due to excessive fussing ($N = 2$), external interference ($N = 3$), and for fatigue ($N = 1$, see Procedures, Experiment 1 for a description of this criterion).

Procedure. The stimulus events and apparatus were identical to those of Experiment 1. Subjects were randomly assigned to the experimental or the control conditions, with 12 in each condition. The 12 events used in the prior study (four composing the set of metal object, four from the set of wooden objects, and four from the set of plastic objects) were randomly assigned to experimental and control subjects for habituation such that one experimental and one control subject received each event. Half the subjects in each condition were presented with the motions of the object in synchrony with the appropriate high-pitched sound and half with the low-pitched sound, during the habituation trials. In all other respects, the habituation procedure was identical to that of Experiment 1. During the two test trials, the experimental subjects received trials of the familiar object moving in synchrony with the novel pitch. Thus, the only change from habituation to test was the pitch of the sound. Subjects in the control condition received no change from habituation to test trials.

A trained observer, unaware of the subject's condition, monitored visual fixations as before. The observer wore Tandy 12-198 radio headphones which produced white noise, masking the stimulus sounds. A second observer monitored fixations for 9 of the 24 subjects for the purpose of calculating interobserver reliability. A Pearson product-moment correlation between the observations of the primary and secondary observers revealed a mean correlation of .99 ($SD = .022$).

Results and Discussion

Visual recovery scores were calculated as before, by subtracting the average fixation time on the two posthabituation trials from that of the two test trials. Subjects in the experimental condition showed a mean visual recovery of 5.31 s (SD = 4.97), while those in the control condition showed a recovery of −2.86 s (SD = 4.77). A t test comparing these means revealed a significant difference ($t(22) = 4.11$, $p = .0005$). Infants were thus able to discriminate between the high- and low-pitched sounds.

Further analyses were conducted to determine whether there were any effects of stimulus set on visual recovery scores. A two-way analysis of variance with condition (experimental, control) and stimulus set (wood, metal, plastic) as between-subjects factors revealed no main effect of stimulus set ($F(2, 18) = .67$, $p > .1$) or interaction of stimulus set with condition ($F(2, 18) = 1.89$, $p > .1$). An analysis was also conducted to determine whether visual recovery scores differed for subjects who had received habituation with high versus low sounds. A two-way analysis of variance, with condition (experimental versus control) and pitch (high versus low) as main factors, revealed no main effect of pitch ($F(1, 20) = 1.01$, $p > .1$) and no interaction of pitch with condition ($F(1, 20) = .122$, $p > .1$).

To determine whether subjects showed any a priori differences in initial interest level, final interest level, or pattern of habituation across the experimental and control groups, t tests were performed separately comparing the performance of experimental and control subjects on the five secondary habituation measures (baseline, number of trials to criterion, number of seconds to criterion, mean criterion fixation time, and mean posthabituation fixation time; see Secondary analyses, Experiment 1, for details). Results indicated no significant differences on any of the measures ($p > .1$, all tests).

In sum, although experimental and control subjects showed no a priori differences in their interest and habituation patterns with respect to the displays, experimental subjects who received a pitch change showed significantly higher visual recovery scores than controls. Infants of 3½ months are able to discriminate between the sounds of the high and low pitches. These findings rule out the alternative interpretation of Experiment 1 that the failure of subjects to detect changes in the pitch–color/shape relations was due to an inability to discriminate the high and low sounds.

EXPERIMENT 3: VISUAL DISCRIMINATION

This study assessed the ability of 3½-month-old infants to discriminate between the events used in Experiment 1 on the basis of their differences in color and shape.

Method

Subjects. Twenty-four subjects, 13 males and 11 females, with a mean age of 112.38 days (SD = 7.5) participated. The data from 12 additional subjects were rejected due to excessive fussing (N = 5), external interference (N = 1), fatigue (N = 3), and failure to habituate (N = 3), (see Procedure Experiment 1 for a full explanation of the fatigue and habituation criteria).

Procedure. The stimulus events and apparatus were identical to those of Experiment 1. Subjects were randomly assigned to the experimental (N = 12) or control (N = 12) conditions. As in Experiment 2, subjects in the experimental and control conditions were randomly assigned to one of the 12 event conditions for the habituation sequence. During habituation, the objects were accompanied by an appropriate and synchronous sound of either a high or low pitch. The pitch of the sound was randomly assigned such that 6 subjects in each condition received the high pitch and 6 received the low pitch. During the test trials, subjects in the experimental condition received only a change in color/shape of the object (identical to the change received by subjects in the pitch–color/shape condition of Experiment 1), while those in the control condition received no change from habituation to test. In all other respects the procedure was identical to that of Experiment 1.

A trained observer, unaware of the subject's condition, monitored visual fixations. A second observer monitored fixations for 8 of the 24 subjects. Interobserver reliability was calculated as before and the correlation between observations of the primary and secondary observer averaged .996 (SD = .003).

Results and Discussion

Visual recovery, from posthabituation to test trials, was calculated as before. Subjects in the experimental condition showed a mean recovery of 12.34 (SD = 11.53), while those in the control condition showed a mean recovery of −1.47 (SD = 4.85). These means were compared with a t test and revealed a significant difference ($t(22) = 3.82$, $p < .002$). Infants of 3½ months were thus able to discriminate the color/shape changes used in Experiment 1.

Further analyses were conducted to test for any effects of stimulus set on visual recovery scores. A two-way analysis of variance was performed with condition (experimental, control) and stimulus set (wood, metal, plastic) as between-subjects factors. Results revealed no main effects of stimulus set ($F(2, 18) = 1.8$, $p > .1$) or interaction of stimulus set and condition ($F(2, 18) = 1.29$, $p > .1$). A further analysis was conducted to determine whether recovery scores differed as a function of whether habituation had been conducted with the sounds of the high versus low one of two side by side films with a soundtrack on the basis of audio–visual synchrony or composition information. Abstraction of amodal relations is apparently promoted by sustained attention to a single, unitary event. This, in turn, may promote acoustically guided visual exploration when that event is later presented in the context of other similar events.

It thus appears that the habituation procedure provides a different and more sensitive index of infants' intermodal functioning than does the intermodal preference method (Spelke, 1976). The intermodal preference procedure may be viewed as a matching task, requiring infants to select which of two films matches a soundtrack they are hearing. In contrast, the habituation procedure may be viewed as a discrimination task, requiring infants to show visual recovery to a change in stimulation. Discrimination of changes in audio–visual relations may precede matching on the basis of these relations. The habituation procedure, by providing infants with only one audio–visual event at a time, is less demanding of infants' attentional capabilities than the intermodal preference method and promotes sustained attention to a single, unitary event, facilitating abstraction of amodal relations.

Finally, this research is unique in that it provided young infants with three audio–visual relations in the same events and then assessed which relations they abstracted. In the natural environment, single events often provide a number of different audio–visual relations. This research makes a first step toward determining which relations are typically abstracted in the course of infants' intermodal exploration of events. It may be that detection of amodal relations guides perceptual learning in early infancy and provides the basis for later learning about modality-specific audio–visual relations.

REFERENCES

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

Bahrick, L. E. (1987). Infants' intermodal perception of two levels of temporal structure in natural events. *Infant Behavior and Development*, **10**, 387–416.

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.

Bertenthal, B. I., Haith, M. M., & Campos, J. J. (1983). The partial-lag design: A method for controlling spontaneous regression in the infant-control habituation paradigm. *Infant Behavior and Development*, **6**, 331–338.

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

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

Horowitz, F., Paden, L., Bhana, K., & Self, P. (1972). An infant-control procedure for studying infant visual fixations. *Developmental Psychology*, **7**, 90.

Kuhl, P. K., & Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior and Development*, **7**, 361–381.

Lawson, K. R. (1980). Spatial and temporal congruity 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 Develoment*, **48**, 820–827.

Reardon, P., & Bushnell, E. W. (1988). Infants' sensitivity to arbitrary pairings of color and taste. *Infant Behavior and Development*, **11**, 245–250.

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

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

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

Spelke, E. S., & Owsley, C. J. (1979). Intermodal exploration and knowledge in infancy. *Infant Behavior and Development*, **2**, 13–27.

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

Walker-Andrews, A. S. (1986). Intermodal perception of expressive behaviors: Relation of eye and voice? *Developmental Psychology*, **22**, 373–377.

Walker-Andrews, A. S., Bahrick, L. E., Raglioni, S. S., & Diaz, I. (1991). Infants' bimodal perception of gender. *Ecological Psychology*, **3**(2), 55–75.

RECEIVED: May 14, 1991; REVISED: October 8, 1991.

pitch. A two-way analysis of variance with condition (experimental, control) and pitch (high, low) as between-subjects factors revealed no main effect of pitch ($F(1, 20) = .22, p > .1$) or interaction of pitch with condition ($F(1, 20) = .95, p > .1$).

Additional analyses were performed to determine whether subjects in the experimental versus the control condition showed any a priori differences in their initial interest level in the stimuli, final interest level, or pattern of habituation. Separate *t* tests were performed on each of the five habituation measures (baseline, trials to criterion, seconds to criterion, mean criterion fixation, mean posthabituation fixation) comparing the performance of experimental subjects with that of control subjects. Results revealed no significant differences in the number of trials to habituation ($t(22) = .54, p > .1$) or the mean fixation time on posthabituation trials ($t(22) = .37, p > .1$). However, subjects in the experimental group showed a significantly greater number of seconds to habituation ($t(22) = 2.43, p < .05$), a larger mean criterion fixation time ($t(22) = 2.16, p < .05$), and a marginally higher baseline fixation level ($t(22) = 1.96, p = .063$) than subjects in the control condition. Thus, although experimental subjects were more interested in the stimuli to begin with and accumulated more seconds of fixation to reach the habituation criterion, their final interest level during the posthabituation trials did not differ from that of the controls. Since the visual recovery scores reflect the change in fixation time to the habituation versus the test displays for a given subject, they should not be affected by differences in initial interest in the habituation displays.

In summary, results demonstrated that 3½-month-old infants were able to discriminate between the stimuli used in Experiment 1 on the basis of a change in color and shape. Experimental subjects who received a change in color/shape from habituation to test showed significantly higher visual recovery scores than control subjects who received no change. These findings rule out the alternative explanation of Experiment 1 that the failure of subjects to detect changes in the pitch–color/shape relations was due to an inability to discriminate between the changes in color and shape.

GENERAL DISCUSSION

In the present research, infants of 3½ months showed significant visual recovery to a change in two amodal relations, audio–visual synchrony and composition relations, but not to a change in the arbitrary relation between color/shape and pitch of a sound, relative to the performance of control subjects. Their failure to detect the change in pitch–color/shape relations was not due to an inability to discriminate the acoustic or visual changes. Infants of this age discriminated all of these visual and acoustic changes in the two subsequent discrimination studies.

These results are consistent with an increasing specificity hypothesis. That is, infants abstract amodal invariant relations developmentally prior to abstracting arbitrary, modality-specific relations. This developmental pattern seems adaptive in that it may promote early learning about the persistent properties and relations characteristic of unitary multimodal events, while delaying learning about less persistent, more variable relations characterizing multimodal events. Audible and visible events typically produce sounds that are in synchrony with their visual impacts, and the sounds and visual impacts of a single element striking a surface are typically shorter and more abrupt than those of a group of elements striking a surface. Detection of these relations may lead to appropriate intermodal knowledge and generalization to properties of similar events. However, the relation between the pitch of a sound and the color/shape of an object is arbitrary and early detection and knowledge of these relations would not generalize appropriately to similar events. Rather, it seems adaptive for these arbitrary relations to be detected only after infants have acquired knowledge about more persistent intermodal relations and are able to effectively use this knowledge to guide intermodal exploration. Intermodal exploration on the basis of amodal relations may in turn lead to intermodal knowledge about arbitrary, artifical relations (Spelke, 1979, 1981).

The present results also confirm those of Bahrick (1988), demonstrating that by the age of 3½ months, infants can detect temporal synchrony relations and audio–visual information specifying the composition of an object when presented with a single film of a moving object and its corresponding soundtrack. To date, this is the youngest age at which infants have been shown to detect bimodal information specifying the composition of an object and the ability to abstract two types of invariant audio–visual relations in a single event.

Apparently, young infants are able to abstract amodal synchrony and composition information in single events prior to the time when they are able to use this information to guide intermodal exploration when more than one event is visible. This conclusion is supported by two lines of evidence. First, Bahrick (1988, discussed earlier) demonstrated that infants who received training with a single film and soundtrack together were later able to use amodal relations to guide exploration, whereas controls who received no training could not. Second, the habituation phase of the present study, like the training phase of Bahrick (1988), also provided infants with the opportunity to abstract two types of amodal relations when presented with a single film and soundtrack together. The significant visual recovery to a change in each type of relation demonstrated their sensitivity to amodal synchrony and composition relations under this procedure. However, an earlier pilot study with the same stimuli demonstrated that infants ($N = 33$) of this age showed no evidence of matching