Developmental Psychology
2005, Vol. 41, No. 3, 541–552

0012-1649/05/$12.00 DOI: 10.1037/0012-1649.41.3.541

The Development of Infant Learning About Specific Face–Voice Relations

Lorraine E. Bahrick
Florida International University

Maria Hernandez-Reif
University of Miami

Ross Flom
Brigham Young University

This study examined the development of infants' ability to perceive, learn, and remember the unique face–voice relations of unfamiliar adults. Infants of 2, 4, and 6 months were habituated to the faces and voices of 2 same-gender adults speaking and then received test trials where the faces and voices were synchronized yet mismatched. Results indicated that 4- and 6-month-olds, but not 2-month-olds, detected the change in face–voice pairings. Two-month-olds did, however, discriminate among the faces and voices in a control study. Results of a subsequent intermodal matching procedure indicated that only the 6-month-olds showed matching and memory for the face–voice relations. These findings suggest that infants' ability to detect the arbitrary relations between specific faces and voices of unfamiliar adults emerges between 2 and 4 months of age, whereas matching and memory for these relations emerges somewhat later, perhaps between 4 and 6 months of age.

Keywords: intersensory perception, intermodal learning, infant perceptual development, face–voice matching, arbitrary intersensory relations

In early infancy, much perceptual, cognitive, and social learning emerges in the context of close face-to-face interaction (e.g., Gibson & Pick, 2000; Rochat, 1999). How do infants make sense of this changing array of multimodal stimulation? For example, how do they learn which faces and voices belong together and constitute unitary events? How do they acquire intermodal knowledge about the people and events in their environment?

Multimodal events make two kinds of information available, amodal and modality specific. Amodal information is not specific to a particular sensory modality but is common across more than one sense (see Bahrick & Lickliter, 2002; Gibson, 1969; Stoffregen & Bardy, 2001; Walker-Andrews, 1994). Most temporal, spatial, and intensity information is amodal, including audiovisual synchrony, tempo, rhythm, colocation, and changing intensity patterns. For example, movement of the lips and the timing of speech share temporal synchrony, rhythm, tempo, and common intensity shifts. By detecting amodal audiovisual relations, perceivers can determine that a face and voice belong together and can identify which of many individuals in a group is speaking. In contrast, multimodal events also make modality-specific information available (information that can be conveyed primarily through one sense modality alone; see Bahrick, 1994; Bahrick & Lickliter, 2002). Examples include pitch and timbre in the auditory domain, color and pattern in the visual domain, and temperature in the tactile domain. For example, each individual is characterized by a voice of a unique pitch and timbre. These properties are primarily experienced acoustically. Individuals are also characterized by a unique facial appearance and configuration of features, properties specific to vision. Thus, audiovisual events provide both amodal relations that unite the sights and sounds of an event as well as arbitrary, modality-specific relations that specify a unique object or individual.

Young infants are excellent perceivers of faces (see Nelson, 2001; Walker-Andrews, 1997, for reviews). Faces are salient, preferred over other stimuli (Easterbrook, Kisilevsky, Hains, & Muir, 1999; Maurer & Barrera, 1981), and easily discriminated in the first months of life (e.g., Fagan, 1972; Kleiner, 1987; Maurer & Young, 1983; Pascalis & de Schonen, 1994). Infants discriminate among static representations of faces including the mother and stranger (e.g., Barrera & Maurer, 1981; Bushnell, 1982; Pascalis, de Haan, Nelson, & de Schonen, 1998; Pascalis, de Schonen, Morton, Deruelle, & Fabre-Grenet, 1995; Walton, Bower, & Bower, 1992) and two strangers (e.g., Cohen & Strauss, 1979; Cornell, 1974; Fagan, 1972, 1976). They also discriminate between moving faces of the mother or father versus a stranger (Carpenter, Teece, Stechler, & Friedman, 1970; Field, Cohen, Garcia, & Greenberg, 1984; Sai & Bushnell, 1988; Spelke & Owsley, 1979), between two strangers (Bahrick, Lickliter, Vaillant, Shuman, & Castellanos, 2004), and between the self and a peer in dynamic and static presentations (Bahrick, Moss, & Fadil,

---

Lorraine E. Bahrick, Department of Psychology, Florida International University; Maria Hernandez-Reif, Department of Pediatrics, University of Miami; Ross Flom, Department of Psychology, Brigham Young University.

This research was supported by National Institute of Child Health and Human Development Grant RO1HD25669 and National Institute of Mental Health Grant RO1 MH62226 awarded to Lorraine E. Bahrick. A portion of these data was presented at the Annual Meeting of the International Society for Developmental Psychobiology, Orlando, Florida, October 2002. We thank Katryna Anasagasti, Laura Batista, Martha Caveda, Maricel Cigales, Melissa Shuman, and Marianna Vaillant for their assistance in data collection.

Correspondence concerning this article should be addressed to Lorraine E. Bahrick, Department of Psychology, Florida International University, Miami, FL 33199. E-mail: bahrick@fiu.edu

1996). Thus, infants perceive and discriminate among faces under a diverse range of conditions in the first months of life.

Young infants are also excellent perceivers of the human voice and the sounds of speech. Fetuses attend to sounds and voices toward the end of gestation (Birnholtz & Benacerraf, 1983; Querleu, Renard, Boutteville, & Crepin, 1989), and infants show prenatal auditory learning of the maternal voice and prosody of speech (e.g., DeCasper & Fifer, 1980; DeCasper & Spence, 1986). Newborns discriminate the voice of their father versus a male stranger and between two male strangers (DeCasper & Prescott, 1984). Young infants also differentiate among voices of strangers (Bahrick, Lickliter, Shuman, Batista, & Grandez, 2003), classify voices on the basis of gender (Miller, 1983; Miller, Younger, & Morse, 1982), and show a preference for infant-directed speech (e.g., Cooper & Aslin, 1990; Fernald, 1984, 1989; Werker & McLeod, 1989). Thus, infants perceive a variety of characteristics of the human voice and patterns of speech in the first months of life.

Typically, in the natural environment, the voices and faces of individuals occur together. They are collocated and share an amodal temporal structure. Young infants are skilled perceivers of amodal information in audiovisual speech and are able to match faces and voices on this basis (see Lewkowicz, 1996; Walker-Andrews, 1997). By the age of 2 months, infants detect voice–lip synchrony (Dodd, 1979), and by 4 months, infants match audible and visible speech on the basis of spectral information in vowel sounds (Kuhl & Meltzoff, 1984; Patterson & Werker, 1999). Infants, like adults, also integrate audible with visible speech, as illustrated by the McGurk effect (Rosenblum, Schmuckler, & Johnson, 1997). When a speech sound is synchronized with the lip movements of a different sound, perceivers often experience a third sound, a combination of the two. Between 5 and 7 months, infants are also able to match faces and voices on the basis of affect (e.g., Soken & Pick, 1992; Walker, 1982; Walker-Andrews, 1997), gender (Walker-Andrews, Bahrick, Raglioni, & Diaz, 1991), and age of speaker (Bahrick, Netto, & Hernandez-Reif, 1998). These diverse findings demonstrate that infants are excellent perceivers of a wide range of amodal relations uniting faces and voices across a variety of events.

Little research, however, has investigated the detection of intersensory relations between modality-specific attributes. Although amodal relations may be perceived directly or with little experience (e.g., a rhythm is the same regardless of whether it is seen or heard), relations between modality-specific properties are arbitrary and must be learned through experience with the event (see Bahrick, 2001; Bahrick & Pickens, 1994; Gibson, 1969). For example, the relation between a particular speech sound "cup" and the appearance of the object it refers to is an arbitrary convention that differs from one language to the next. Similarly, the relation between the sight of a person's face and the unique sound of his or her voice, or the particular ringing sound and the telephone it belongs with, are arbitrary. Perceivers must acquire intermodal knowledge about these arbitrary modality-specific relations through experience with the particular events.

Research indicates that infants detect modality-specific relations in multimodal stimulation but that detection of these relations emerges somewhat later than detection of amodal relations in the same events (see Bahrick, 1992, 1994, 2001). Detection of multimodal relations appears to progress in order of increasing specificity, from amodal relations that are considered more global to modality-specific relations (e.g., Bahrick, 1992, 1994, 2001; Bahrick & Lickliter, 2002; Gibson, 1969; Walker-Andrews, 1997). Further, detection of amodal information guides and constrains perceptual learning about arbitrary, modality-specific relations (Bahrick, 1988, 1992, 1994, 2001, 2004; Gogate & Bahrick, 1998; Hernandez-Reif & Bahrick, 2001; Slater, Quinn, Brown, & Hayes, 1999). Detection of arbitrary relations has been shown to occur in the presence but not in the absence of amodal temporal synchrony. For example, at 7 months, infants learned to pair a particular speech sound with an object, only when temporal synchrony united the movements of the object with the sounds and not in the absence of synchrony or when the object was still (Gogate & Bahrick, 1998). Similarly, 7-month-old infants, but not 3- or 5-month-old infants, detected the arbitrary relation between the color–shape of an object and the pitch of its impact sounds when the sounds and sights of impacts were synchronous (Bahrick, 1994). In contrast, 3-month-olds detected the amodal temporal relations available in these events (Bahrick, 1992). Three-month-olds also showed generalization of learning on the basis of these amodal temporal relations to objects that were somewhat different (new color and shape) but not to objects that were very different (new color, shape, substance, and type of motion) from those of original learning (Bahrick, 2002). At 3 months, infants also showed better detection and memory for the arbitrary relation between the visual appearance of an object and a collocated rattling sound in the presence of object–sound synchrony than in its absence. By the age of 7 months, infants no longer required synchrony for detection and memory of the collocated object–sound pairings (Morrongiello, Lasenby, & Lee, 2003). Even newborns have been shown to detect arbitrary audiovisual sound–sight relations in the presence of amodal information (contingent presentations) but not in the absence of amodal information (noncontingent presentations; Slater, Quinn, Brown, & Hayes, 1999). Thus, amodal relations provide a basis for guiding attention and constraining learning to unitary audiovisual events and subsequently for appropriate generalization of knowledge across a variety of audiovisual events. Further, some arbitrary object–sound relations can apparently be detected even by young infants and neonates, as long as the audiovisual presentations are synchronous or contingent with the infant's motions.

The intersensory redundancy hypothesis (see Bahrick & Lickliter, 2000, 2002) makes complementary predictions about the course of infants' detection of specific face and voice information as a function of whether stimulation is perceived unimodally or bimodally. According to the intersensory redundancy hypothesis, bimodal stimulation is highly salient and provides redundantly specified amodal properties such as voice–lip synchrony, tempo, rhythm, and prosody of speech. Experiencing an event redundantly across two senses facilitates perception of amodal properties, whereas experiencing an event in one sense modality alone (with no redundancy) facilitates perception of modality-specific properties of stimulation such as pitch and timbre of the voice or pattern and configuration of the face (see Bahrick, Flom, & Lickliter, 2002; Bahrick & Lickliter, 2000, 2002; Bahrick, Lickliter, & Flom, 2004). Therefore, in early development, discrimination of faces or voices (which relies on modality-specific information) is enhanced when faces or voices are explored unimodally or without redundancy and attenuated when they are explored bimodally with redundancy (Bahrick et al., 2003; Bahrick, Lickliter, Vaillant, et al., 2004). For example, 2-month-olds showed unimodal discrimination of faces (during silent visual speech) but not bimodal discrimination (during audiovisual speech; Bahrick, Lickliter, Vaillant, et al., 2004). Thus, in naturalistic events, optimal conditions for very young infants to perceive the appearance of faces and the specific sound of voices are when a person is seen but not heard (e.g., silent phases of face-to-face interaction) and when the person is heard speaking but not seen (e.g., the infant's eyes are closed or gaze is averted during speech; the speaker is heard from a nearby room). This is because when attention is relatively inflexible and the task is difficult for infants, in bimodal stimulation (when the face and voice are experienced together), amodal properties such as synchrony or prosody are most salient and attended first. Later in development, as attention becomes more flexible and less constrained, infants can detect both amodal and modality-specific properties in bimodal and unimodal stimulation (Bahrick & Lickliter, 2004; Bahrick, Lickliter, & Flom, in press), and by 3 months of age, infants can discriminate among faces even in bimodal audiovisual speech (Bahrick & Lickliter, 2002; Bahrick, Lickliter, & Flom, 2004). Thus, there is a developmental shift in the perception of multimodal events, from detection of amodal information to detection of amodal and modality-specific information for a given event as a function of experience.

The developmental principles described above provide an account of how infants can learn to detect appropriate, arbitrary, audiovisual relations between the faces and voices of individuals, without learning inappropriate audiovisual relations (such as pairing the voice of one person with a nearby object or face of a different person). Detection of amodal information (e.g., temporal synchrony, rhythm, and tempo) promotes attention to a unitary event, the speaking person. Further differentiation of the unitary event promotes attention to increasingly more specific aspects of stimulation, eventually leading to learning about the arbitrary relations such as those between the pitch and/or timbre of the voice and the specific configuration of the face. This process promotes intermodal knowledge and appropriate generalizations about audiovisual relations that reliably belong together (see Bahrick, 2001, 2004). Thus, the redundancy provided by amodal information facilitates detection of unitary face–voice events and constrains learning about specific faces and voices until unitary events and global, amodal properties are perceived.

Although it is clear that faces and voices are salient and discriminated in early infancy, it is not clear when and under what conditions infants learn to relate particular faces and voices of individuals and develop intermodal knowledge about these unique face–voice relations. Few studies have investigated matching of arbitrary face–voice pairings based on modality-specific attributes. One study (Spelke & Owsley, 1979) demonstrated that 4-month-olds could match the faces and voices of very familiar adults, their mother and father, even when voice–lip synchrony was controlled. However, it is not clear whether infants detected the arbitrary and unique face–voice relations or whether typical and/or amodal relations specifying gender (e.g., men have larger features and voices of a lower pitch) played a role. Four-month-olds detected intermodal relations specifying gender, even in films of unfamiliar adults (Walker-Andrews et al., 1991).

Only one study to date has investigated infants' perception of arbitrary face–voice relations for unfamiliar adults (Brookes et al., 2001). Three-month-old infants were familiarized with videos of two individuals speaking (either a male and female speaker or two individuals of the same gender). On test trials, the familiarized faces and voices were mismatched but synchronized. Infants showed robust evidence of detecting the mismatch in face–voice relations when the mismatch portrayed a gender violation (the male face was synchronized with the female voice and vice versa), converging with findings of Walker-Andrews et al. (1991). However, in the Brookes et al. (2001) study, infants showed only weak evidence of detecting the mismatch in face–voice relations when faces and voices were of the same gender (one-tailed effects; and 11 of 16 infants showed a novelty preference, a nonsignificant effect). Matching faces and voices within gender appears to be more difficult than matching across gender. These results suggest that by 3 months of age, infants may be just beginning to perceive and learn arbitrary intermodal relations between the unique appearance of the face and specific sounds of the voice of unfamiliar individuals.

The present research examined the developmental course of infants' ability to detect, learn, and remember the arbitrary face–voice relations of two adult strangers of the same gender. It extended the research of Brookes et al. (2001) in a number of ways. The emergence and development of this ability was investigated across age, the research focused on infants' detection of arbitrary audiovisual relations and thus only matching of same-gender adults was assessed, and we also assessed memory for the face–voice relations.

Experiment 1

Infants of 2, 4, and 6 months were habituated in an infant-controlled procedure with two face–voice pairings of same-gender adults (e.g., Shirley's face and voice vs. Margie's face and voice) across alternating trials and then received test trials where the faces and voices were mismatched (e.g., Shirley's face with Margie's voice and Margie's face with Shirley's voice). Visual recovery to the change reflected learning of the arbitrary face–voice relations. After a 10-min break, infants of 4 and 6 months were also tested for their ability to match and remember the face–voice relations in an intermodal matching procedure. Infants viewed the two faces side by side, along with the centrally projected voice of one, synchronized with the movements of both. Because voice–lip synchrony and sound localization were controlled, matching of the faces and voices required learning and memory for the arbitrary face–voice relations perceived during habituation.

Method

Participants

Twenty 2-month-olds (9 girls and 11 boys) whose mean age was 65 days (*SD* = 6.0), 20 4-month-olds (14 girls and 6 boys) whose mean age was 122 days (*SD* = 6.7), and 20 6-month-olds (8 girls and 12 boys) whose mean age was 186 days (*SD* = 11.7) participated. Sixty-seven percent of the participants were Hispanic, 18% were White not of Hispanic origin, 13% were of African American origin, and 2% were of Asian origin. All were healthy, normal, full-term infants weighing at least 5 pounds at birth, with 5-min Apgar scores of 9 or higher. The data of 15 infants at 6 months, 12 infants at 4 months, and 35 infants at 2 months were rejected from the study, suggesting that the task was rather demanding for the 2-month-olds.

At 6 months, 6 infants were rejected for excessive fussiness, 5 for equipment failure, and 4 for experimenter error. At 4 months, 5 infants were rejected for excessive fussiness, 2 for experimenter errors, and 5 for excessive side bias during the intermodal matching test (see *Procedure* section for details). At 2 months, 18 were rejected for excessive fussiness, 3 for experimenter error, 2 for equipment failure, 4 for failing to habituate within 20 trials, and 8 for failure to pass the criteria for fatigue ($n = 5$) or attention ($n = 3$; see *Procedure* section for details). Parents of the participants were initially contacted by telephone and/or bulk mailings.

Stimulus Events

Color video films of two male and two female adults speaking a nursery rhyme were taken from films used in a prior study (Walker-Andrews et al., 1991). The actors ranged in age from 25–30 years, all were Caucasian, and one member of each pair had light hair and the other had dark hair. The films depicted the face and shoulder area while the actor spoke the nursery rhyme "This Old Man." Synchrony across and between the actors' audio and visual speech was accomplished by having the actors practice speaking the nursery rhyme in synchrony with the lip movements of a videotaped adult model (see Walker-Andrews et al., 1991). After recording the video portion, all actors dubbed the speech onto the video portion of their own face with careful attention to voice–lip synchrony. The actors also incorporated the facial expressions and affect of the model, specifically the timing of the smiles and eyebrow movements at the beginning of each verse, in order to equate for affect and overall amount of motion across actors. This allowed presentation of any two events side by side with excellent voice–lip synchrony and similar changes in affect and facial movements. In addition, a control stimulus was used depicting a plastic green and white turtle whose front legs spun and produced a whirring sound.

Apparatus

The stimulus events were videotaped with a Panasonic (WV 3170) color video camera and a Sony (EMC 150T) remote microphone. The events were edited and presented with a Panasonic (VHS NV-A500) edit controller that was connected to four Panasonic video decks (AG-6300 and AG 7500). The video decks were connected to two 19-in. (48-cm) color video monitors (Sony KV-20M10). Four video decks allowed us to switch between the habituation, test, and control displays without extra time or noise that would have resulted from changing videocassettes. Soundtracks were presented from a speaker located between the monitors at approximately 65 dB, as measured from the infant seat. During the intermodal matching phase, the video displays were presented using two video decks that were connected to the edit controller. The edit controller enabled precise synchronization (to the nearest frame) of the output from the two video decks so that the lip movements of the actors were aligned.

Infants sat facing the video monitors approximately 50 cm away. Two apertures cut into the black cloth surrounding the monitors were used to record the infants' visual fixations. The observers, unaware of the hypotheses of the experiment and unable to view the visual events, depressed a button while the infant fixated on the event and released it while the infant looked away. They also wore headphones that played music so they would be blind to the auditory information presented to the infant. The observers' button boxes were connected to a computer programmed to record visual fixations online. The computer signaled through a small earphone to a second experimenter who controlled the presentation of the video displays when to end the trials and when the habituation criterion had been reached. The observations of the primary observer controlled the audiovisual presentations, and those of the secondary observer were used in the computation of interobserver reliability.

Procedure

All procedures were approved by the institutional review board for the protection of participants at Florida International University. Infants first received infant-controlled habituation trials (Horowitz, Paden, Bhana, & Self, 1972) to two face–voice pairings presented on alternating trials, and detection of these arbitrary pairings was assessed by visual recovery to mismatched test trials. The same procedure as that used in our prior studies was followed (e.g., Bahrick, 1992, 1994, 2001). Then, following a 10-min break, the 4- and 6-month-olds participated in the second phase, an intermodal matching procedure. This assessed whether infants had learned and could remember the face–voice relations of the habituation phase by selecting one of the two faces that belonged with the voice heard. (Two-month-olds did not participate because too many were rejected for fussiness during the habituation phase.)

Within each age group, infants were randomly assigned to receive the two male or two female events (half in each condition). Within each of these groups, half of the infants were habituated with faces that were paired with their own voices (e.g., Margie's face with Margie's voice and Shirley's face with Shirley's voice), and the other half were habituated with faces paired with the other voice (e.g., Margie's face with Shirley's voice and Shirley's face with Margie's voice).

Habituation and recovery. Each habituation sequence consisted of at least six infant-controlled habituation trials (three of each face–voice pair) and was terminated once the infant reached the habituation criterion and completed two subsequent no-change posthabituation trials (one for each face–voice pair). A trial began when the infant looked toward the video display and ended when the infant looked away for greater than 1.5 s. Sixty seconds was set as the maximum trial length, and 20 trials was the maximum number of trials. The habituation criterion was defined as a 50% decline in looking, on two consecutive trials, compared with the infant's average looking time on the first two trials (i.e., baseline trials). After the habituation criterion was met, the two no-change posthabituation trials were presented. These additional habituation trials were included to establish a more conservative habituation criterion by reducing the possibility of chance habituation and taking into account spontaneous regression toward the mean (see Bertenthal, Haith, & Campos, 1983, for discussion of regression effects in habituation designs). These trials also served as a basis for assessing visual recovery. Following the two no-change posthabituation trials, infants received two infant-controlled test trials where the face–voice pairings were switched (e.g., from Margie's face with Shirley's voice and Shirley's face with Margie's voice, to Margie's face with Margie's voice and Shirley's face with Shirley's voice) to assess detection of the arbitrary face–voice relations. Prior to beginning the habituation sequence, we presented the control event (toy turtle) as a warm-up trial, and it was also used after the presentation of the test trials to examine infants' level of fatigue.

Exclusion criteria. We examined each infant's data to determine whether two criteria had been met (see Bahrick, 1992, 1994, 2001). First, to ensure that infants actually habituated, we compared the infant's mean posthabituation looking to their initial baseline looking. If an infant's mean posthabituation looking was equal to or greater than their initial looking (baseline), (a) this indicated the habituation trials had no effect on the infant's final level of looking, (b) it was judged the infant had not habituated, and (c) their data were excluded from the analyses. The data of three 2-month-olds were excluded. Second, to exclude the data of infants who were overly fatigued and unable to show visual recovery, we compared infants' looking on the first and final control trials (i.e., the moving turtle). On the final control trial, infants were required to look at least 20% of their initial looking level. The data of five 2-month-olds were excluded for failure to meet this criterion.

Intermodal matching. The 4- and 6-month-olds received a 10-min break before beginning the intermodal matching phase. The mother typically held the infant in the testing room and then the infant was again seated and positioned equidistant from two side-by-side monitors. Only infants who successfully completed the habituation phase were included in the intermodal matching phase. During the intermodal matching phase, infants were presented with two identical blocks of six 15-s trials. Each trial depicted the two adults seen during habituation speaking in synchrony side by side, along with the synchronized voice matching one of the two faces. In each block, each voice was played on three trials. The ordering of voices within each block was randomized, with the restriction that the same voice did not occur on more than two consecutive trials. The second block was identical to the first block, except the right–left positions of the faces changed from the first to the second block. Intertrial and interblock intervals were approximately 3 s long. Trained observers who were unaware of the lateral positions of the video displays monitored the infants' looking times.

Exclusion criteria. Two looking criteria were set for the intermodal matching data of the participants to be included in the analyses. Infants were required to complete five out of six trials for each block. No data were rejected for failure to meet this criterion. We also felt that it was important that infants notice that there were two video events side by side. Thus, an attention criterion required that infants look at least 5% of the time to the least preferred display for the trial to be included. If there were not five of six usable trials remaining, the data of that infant were rejected. The data of five 4-month-olds were rejected for failure to meet this criterion.

Results

Habituation Phase

A secondary observer monitored the visual fixations for 35% of the 2-month-olds, 20% of the 4-month-olds, and 20% of the 6-month-olds. Interobserver reliability was calculated by correlating the visual fixation scores of the primary and secondary observers across trials for each infant and it averaged .98 ($SD = .02$) across infants in the three age groups. An analysis of variance (ANOVA) was first performed, with gender of the infants as a main factor, and revealed no significant effect of gender for any of the trial types at any age (all $ps > .10$). Thus, all analyses were collapsed across gender.

Figure 1 depicts the mean visual fixation as a function of trial type (baseline, posthabituation, and test) for infants of each age. An overall repeated measures ANOVA, with age (2, 4, and 6 months) as a between-subjects factor and trial type (baseline, posthabituation, and test) as the repeated measure, was first performed to compare looking behavior across age and trials. All significance levels are reported with two-tailed values. Results revealed a significant main effect of trials, $F(2, 114) = 112.40$, $p < .001$, η^2_p (partial effect size) = .66, a significant main effect of age, $F(2, 57) = 20.60$, $p < .001$, $\eta^2_p = .42$, and a significant Trials × Age interaction, $F(4, 114) = 3.06$, $p = .02$, $\eta^2_p = .10$. Planned comparisons explored the main effect of age and revealed that the overall looking performance of the 6-month-olds differed from that of the 4-month-olds ($p < .001$) and from that of the 2-month-olds ($p < .001$) but that the overall looking of the 2- and 4-month-olds did not differ from one another ($p > .10$). Thus, as can be seen in Figure 1, the 6-month-olds showed less looking overall, across trial types, than the 2- and 4-month-olds. Analyses of simple effects revealed a main effect of age for each variable, baseline, posthabituation, and test, $F(2, 59) = 12.60$, $p < .001$;

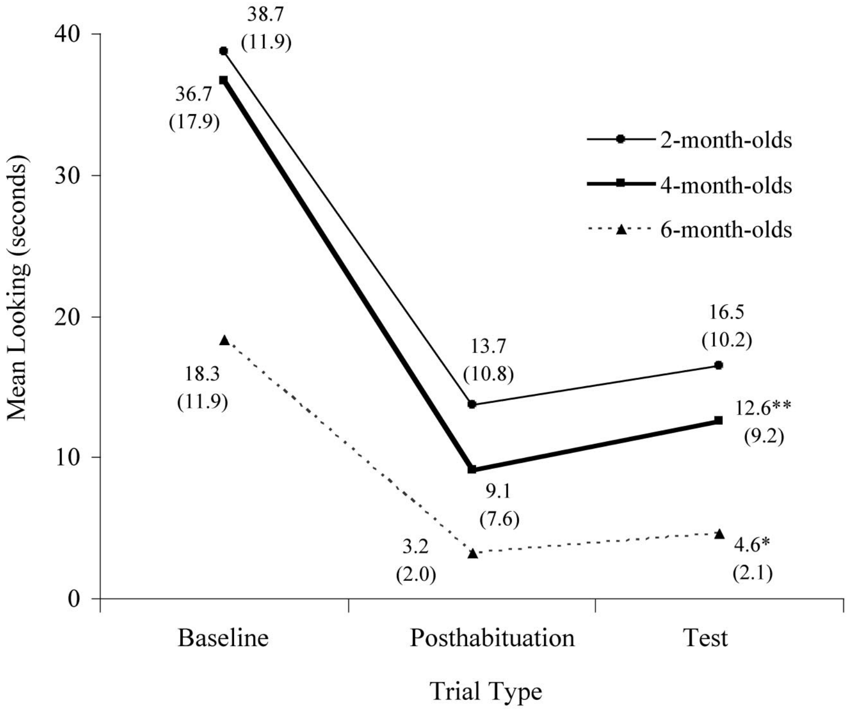

Figure 1. Mean visual fixation (and standard deviations) as a function of trial type at 2, 4, and 6 months of age during the habituation phase. Baseline is the mean visual fixation during the first two habituation trials and reflects initial interest. Posthabituation is the mean visual fixation to two no-change trials just after the habituation criterion was met and reflects final interest in the habituated event. Test is the mean visual fixation during the two face–voice change trials. * $p < .05$. ** $p < .01$.

$F(2, 59) = 9.49$, $p < .001$; $F(2, 59) = 11.46$, $p < .001$, respectively, with significantly greater initial interest, final interest, and test trial interest by 2- and 4-month-olds than by 6-month-olds ($p < .05$ for all post hoc differences). Thus, for all trial types, younger infants spent more time observing the visual displays than the older infants. Finally, we also conducted an ANOVA, with age as the main factor, on the number of seconds required to reach habituation. Results also revealed a significant main effect of age, $F(2, 59) = 10.11$, $p < .001$, with 2-month-olds ($M = 239.6$, $SD = 177.4$) showing no difference from that of the 4-month-olds ($M = 228.1$, $SD = 126.4$) in time to reach habituation, and both the 2- and the 4-month-olds taking longer to habituate than the 6-month-olds ($M = 75.8$, $SD = 45.8$). Thus, for all variables, the 2- and 4-month-olds showed more looking to the faces and voices than the older infants. These results are consistent with prior findings that younger infants require more time to process or explore stimuli than older infants (Fagan, 1974; Hale, 1990; Rose, 1983; Rose, Feldman, & Jankowski, 2002).

Planned comparisons exploring the main effect of trials (baseline, posthabituation, and test) were also conducted and indicated that mean looking for each type of trial differed significantly from that of each other trial type (all ps $< .001$), with greatest looking during baseline and least looking during posthabituation. Thus, across all ages combined, infants showed evidence of habituation (baseline vs. posthabituation trials) and evidence of visual recovery to a change in face–voice relations by a significant increase in looking from posthabituation to test trial performance.

To address the main research question—At what ages did infants detect the change in face–voice relations?—we evaluated the nature of the Trials × Age interaction. Looking patterns across trials were evaluated by one-way repeated measures ANOVAs for each age separately. Then, evidence of infants' detection of face–voice relations was assessed by planned comparisons across trials at each age. Given that infants had habituated to the face–voice pairings, it was expected that infants at all ages would show a significant decrease in looking from baseline to posthabituation. However, only groups of infants who detected the change in face–voice relations should show evidence of visual recovery to the change, reflected by a significant increase in looking from posthabituation to test trial looking. Results of the repeated measures ANOVAs on trial type (baseline, posthabituation, and test) for the 2-, 4-, and 6-month-olds separately revealed a significant main effect of trials at each age, $F(2, 38) = 38.44$, $p < .001$; $F(2, 38) = 43.89$, $p < .001$; $F(2, 38) = 31.7$, $p < .001$, respectively. The trials effects were analyzed by planned comparisons. As expected, infants of all ages demonstrated a significant decrease in looking from baseline to posthabituation (ps $< .001$ for all ages). However, only the 4- and 6-month-olds showed a significant increase in looking from the posthabituation to the test trials ($p = .007$ and $p = .015$, respectively), demonstrating visual recovery to the change in face–voice pairings. Two-month-olds failed to show significant visual recovery to the change ($p > .10$). Given that both the faces and the voices were familiar, and only the relation between them changed from habituation to test, this visual recovery reflects the infants' ability to discriminate a change in the pairings of the faces and voices. Thus, these findings indicate that 4- and 6-month-old infants, but not 2-month-olds, detected the change in the arbitrary face–voice relations from habituation to test.

It is interesting that the mean visual recovery (difference between mean visual fixation on test trials vs. posthabituation trials) did not increase across age as sensitivity to face–voice relations emerged (see Figure 1). Rather, variability appeared to decrease across age. A Levene test of the equality of variance was conducted on visual recovery, with age as a factor, and revealed a significant effect, $F(1, 38) = 11.49$, $p = .002$. Thus, variance decreased significantly from 2 to 6 months of age, as infants' skill in detecting face–voice relations improved.

We also conducted secondary analyses to assess whether the gender of the faces and voices or pairing of the faces and voices (face–voice pair from same speaker vs. different speakers) used during habituation affected infants' visual recovery to the change in the face–voice relationship. The results of a three-way ANOVA on visual recovery, with age (2, 4, and 6 months), face–voice pairing (same, different), and gender of the actors as between-subjects factors, indicated no main effects of age, pairing, gender, or interactions of these variables on visual recovery to the change in face–voice relations (all $ps > .10$). Thus, neither gender of the actors nor face–voice pairing affected infants' ability to detect a change in the arbitrary face–voice relationship at any age.

Intermodal Matching Phase

The purpose of the intermodal matching phase, given only to the 4- and 6-month-olds, was to examine whether infants had learned and could remember the face–voice pairings they had explored during habituation. If so, we expected them to look longer in a two-choice intermodal matching procedure to the face that had previously been paired with the voice they were hearing. The results of the intermodal matching phase were expressed in terms of the proportion of total looking time (PTLT) infants looked to the voice-matched face. Proportions were derived for each trial separately by dividing the time spent looking to the voice-specified face by the time spent looking at both faces. These proportions were then averaged to obtain the mean proportion for Block 1 (Trials 1–6) and Block 2 (Trials 7–12) for each infant and across all infants. An overall PTLT was also derived by averaging across the two blocks for each infant and then averaging over all infants (see Figure 2). ANOVAs, with gender of the infant as the main factor, were performed on each measure and revealed no significant effects at either age (ps $> .10$). Thus, subsequent analyses were collapsed across gender. A secondary observer recorded looking times for 30% of the infants at each age. Interobserver reliability was expressed as a Pearson product–moment correlation between the looking proportions of the primary and secondary observers and was .91 ($SD = .02$), averaged across infants of both ages.

In order to determine whether the 4- and 6-month-olds demonstrated intermodal matching of the faces and voices on the basis of learning during the habituation phase, we compared the mean PTLTs against the chance value of .50 (an equivalent proportion of time spent looking toward each display) at each age. Results indicated that the 4-month-olds failed to demonstrate face–voice matching either on Block 1, Block 2, or across Blocks 1 and 2 combined (ps $> .10$, $\eta_p^2 = .30$, $\eta_p^2 = .10$, $\eta_p^2 = .27$, respectively). The results of the 6-month-olds, however, revealed significant evidence of face–voice matching on Block 1, $t(19) = 2.15$, $p =$

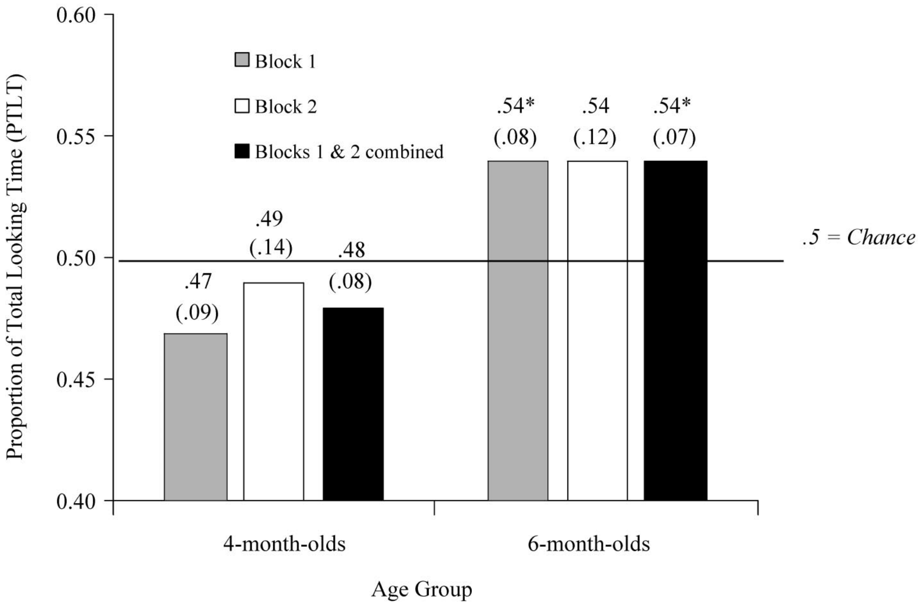

Figure 2. Mean proportion of total looking time (and standard deviations) to the voice-matched face for Trial Block 1, Trial Block 2, and Trial Blocks 1 and 2 combined as a function of age during the intermodal preference phase. * $p < .05$.

.044, $\eta_p^2 = .44$, and on Blocks 1 and 2 combined, $t(19) = 2.42$, $p = .03$, $\eta_p^2 = .49$ (see Figure 2). The results of Block 2, however, failed to reach significance for the 6-month-olds ($p > .10$, $\eta_p^2 = .33$). It may be that once the 6-month-olds detected the face–voice relations in the first block, they became less interested in the task by the second block of trials (see Bahrick et al., 1996, 1998; Walker-Andrews et al., 1991, for similar results across trial blocks). Thus, these results indicate that although both 4- and 6-month-olds are able to perceive and discriminate a change in arbitrary face–voice relations in this procedure, only the 6-month-olds show evidence of remembering the face–voice relations and matching a voice they are hearing with the appropriate face.

In order to evaluate the change in face–voice matching across age, we conducted separate ANOVAs, with age as the between-subjects factor, on the PTLT to the voice-specified face for each block of trials and for both blocks combined. The PTLT of the 6-month-olds was significantly greater than the mean PTLT of the 4-month-olds, for Block 1 and for Blocks 1 and 2 combined, $F(1, 38) = 6.15$, $p = .02$, $\eta_p^2 = .37$; $F(1, 38) = 6.20$, $p < .02$, $\eta_p^2 = .37$, respectively. For Block 2, the mean PTLT for the 6-month-olds and 4-month-olds did not differ ($p > .10$, $\eta_p^2 = .21$). Thus, on the basis of a 3–5-min exposure to arbitrary face–voice relations of unfamiliar adults, infants showed a significant increase in memory and face–voice matching between the ages of 4 and 6 months.

Further analyses were conducted to explore the failure of the 4-month-olds to show intermodal matching of the face–voice relations. To evaluate whether 4-month-olds were fatigued given their length of looking during the habituation phase ($M = 228.12$ s, $SD = 126.4$) and therefore unable to show matching, we compared the performance of fast and slow habituators (defined by a median split; see Colombo, Mitchell, Coldren, & Freeseman, 1991; Freeseman, Colombo, & Coldren, 1993; Frick & Colombo, 1996). The fast habituators ($M = 126.0$ s, $SD = 59.1$) and slow habituators ($M = 330.2$ s, $SD = 84.3$) were evaluated separately in terms of their performance during the intermodal matching phase. If fatigue was a factor in their intermodal matching, then it would be expected that slow habituators (those who spent more time looking during habituation) would be more fatigued and would perform poorly relative to fast habituators. One-way ANOVAs were performed on matching during Block 1, Block 2, and Blocks 1 and 2 combined, with habituation type (fast vs. slow) as the main factor, and indicated no differences between fast and slow habituators on any of the matching measures (all ps $> .10$). Thus, it is unlikely that fatigue was a factor in the matching performance. An interference explanation for the failure of 4-month-olds to show intermodal matching was also evaluated. The 4-month-olds spent longer looking at the films during both habituation and test than the 6-month-olds, and although they spent longer viewing the matching faces and voices, they also viewed the mismatching faces and voices longer. In order to address the possibility of interference from viewing mismatching faces and voices, we classified the 4-month-olds as having positive visual recovery scores ($n = 12$) versus 0 or negative visual recovery scores ($n = 8$), and the performance of each group was evaluated on the intermodal matching test. If interference was a factor, then infants who showed greater visual recovery should perform more poorly on the intermodal matching test. Results, however, were in the direction opposite that predicted by the interference hypothesis, with infants who had high visual recovery showing significantly greater matching than those with low visual recovery on Block 2, $F(1, 18) = 4.35$, $p = .05$, and marginally on Blocks 1 and 2 combined, $F(1, 18) = 4.01$, $p = .06$. Thus, there was no evidence of interference from the test trials of the habituation phase on intermodal matching.

Discussion

The results of the habituation phase of this experiment demonstrate that 4- and 6-month-olds, but not 2-month-olds, discriminated a change in the arbitrary face–voice relations of unfamiliar adults. However, only 6-month-olds matched the voices with the appropriate one of two faces after a 10-min delay. The 6-month-olds were able to use intermodal knowledge acquired during the habituation phase to guide their audiovisual exploration of faces and voices during the intermodal matching task. Four-month-olds, however, showed no evidence of this ability, and two alternative hypotheses, fatigue and interference from viewing mismatching faces and voices during the test trials, were ruled out as likely explanations. This developmental shift from early discrimination of audiovisual changes (indexed by the habituation procedure) to later intermodal matching on the basis of these changes is consistent with prior findings where discrimination also preceded matching (see Bahrick & Pickens, 1994; Walker-Andrews, 1997).

Further, results of the habituation phase revealed a developmental shift between 2 and 4 months of age in infants' detection of arbitrary face–voice relations. Following a 3–5-min exposure to the events, 4-month-olds', but not 2-month-olds', discrimination of changes in face–voice relations was apparent. At least two potential interpretations exist for the failure of the 2-month-olds to detect changes in face–voice relations. One possibility is that infants simply failed to notice or remember the arbitrary relations between the faces and voices. A second possibility is that 2-month-olds were unable to discriminate the individual faces and/or voices (i.e., failure of unimodal discrimination). Although it is unlikely that 2-month-olds were unable to discriminate the moving faces or the voices given previous research showing voice discrimination (see DeCasper & Fifer, 1980; DeCasper & Prescott, 1984) and face discrimination in young infants (see Nelson, 2001; Walker-Andrews, 1997, for reviews), we nevertheless tested this possibility in Experiment 2. If 2-month-olds are capable of discriminating the unimodal faces and voices, then it would seem more likely that their failure in Experiment 1 stems from an inability to notice or remember the arbitrary intermodal relationship between them.

Experiment 2

Experiment 2 examined 2-month-olds' ability to discriminate a change in the faces or voices used in the previous experiment. In the face-change condition, infants were habituated to one of the silent moving faces and received test trials with the novel moving face of the same gender. In the voice-change condition, infants were habituated to one of the speaking voices and received test trials with the novel voice of the same gender. Visual recovery to the unimodal change in face or voice served as the measure of discrimination. It was expected that the 2-month-olds would discriminate the unimodal changes and that there would be no difference between the unimodal visual and unimodal auditory discrimination.

Method

Participants

Sixteen 2-month-olds (9 girls and 7 boys) whose mean age was 71 days ($SD = 5.6$) participated. Ninety-four percent were Hispanic and 6% were White, not of Hispanic origin. The data of 15 additional infants were rejected, 4 for experimenter error, 1 for equipment failure, 4 for fussiness, 3 for failure to habituate within 20 trials, and 3 for fatigue.

Stimulus Events and Apparatus

Infants received either the visual (silent speech) or auditory portions of the speech events used in Experiment 1. The soundtracks were played with a static image of the associated face taken from each of the videos used in Experiment 1. The apparatus was the same as that used in Experiment 1.

Procedure

Sixteen 2-month-olds were randomly assigned to either the unimodal visual face-change condition or the unimodal auditory voice-change condition. All habituation and counterbalancing procedures were identical to those of the previous experiment, with the exception that infants were only habituated to one face or one voice. In the unimodal visual condition, infants were habituated to one of the male or female faces silently speaking the nursery rhyme. Following habituation and the two no-change posthabituation trials, infants received two test trials with the novel face of the same gender silently speaking. In the unimodal auditory condition, infants were habituated to one of the male or female voices speaking the nursery rhyme while they viewed a static image of the face of that adult. Presentation of the voice was infant controlled and contingent on looking to the static face (see Bahrick & Lickliter, 2000; Walker-Andrews & Grolnick, 1983; Walker-Andrews & Lennon, 1991, for a similar auditory discrimination procedure). Following habituation and the two no-change posthabituation trials, infants received two test trials where a novel voice of the same gender was played speaking the nursery rhyme along with the same static face.

Results and Discussion

The mean amount of time infants spent looking to the novel face or novel voice as a function of trial type (baseline, posthabituation, and test) as well as seconds to habituation and visual recovery are presented in Table 1 along with the data from the 2-month-olds of

Table 1
Means and Standard Deviations for Visual Fixation for 2-Month-Olds for Baseline, Posthabituation, Test Trials, Seconds to Habituation, and Visual Recovery for Multimodal Changes (Experiment 1) Versus Unimodal Changes (Experiment 2) During the Habituation Phases

	Experiment 2: Unimodal		Experiment 1: Multimodal	
Variable	*M*	*SD*	*M*	*SD*
Baseline	36.9	10.4	38.7	11.9
Posthabituation	10.0	9.6	13.7	10.8
Test	19.3	12.7	16.5	10.2
Seconds to habituation	285.4	164.4	239.6	177.4
Visual recovery	9.3	11.5	2.8	7.8

Note. Baseline is the mean visual fixation during the first two habituation trials and reflects initial interest. Posthabituation is the mean visual fixation to two no-change trials just after the habituation criterion was met and reflects final interest in the habituated events. Test is the mean visual fixation during the two change test trials, and visual recovery is the difference between visual fixation during the test trials and visual fixation during the posthabituation trials.

Experiment 1, where habituation was bimodal. We initially compared the performance of the infants in the unimodal auditory and unimodal visual conditions of Experiment 2 on five measures of habituation: (a) baseline (average length of the first two habituation trials), (b) mean number of trials required to reach habituation, (c) mean number of seconds to habituation, (d) mean length of fixation on the two habituation criterion trials, and (e) mean length of fixation on the two no-change posthabituation trials. One-way ANOVAs were conducted for each of these measures to determine whether the unimodal auditory and unimodal visual conditions differed. Results indicate no significant differences for any of the measures (ps $>$ .10). An ANOVA was also conducted on visual recovery, with condition (unimodal auditory vs. unimodal visual) and gender of the infant as main factors, and indicated no significant effects or interaction (ps $>$ .10). Therefore neither infants' patterns of habituation nor their visual recovery differed for the unimodal auditory versus unimodal visual conditions, and the two unimodal conditions were collapsed for all subsequent analyses. Further, given no gender effects, the data were also collapsed across infant gender. Interobserver reliability was calculated as before on 30% of the infants and averaged .98 (SD = .03).

To address the main research question, we conducted a repeated measures ANOVA on trial type (baseline, posthabituation, and test) and followed by planned comparisons examining visual recovery, the difference between mean posthabituation and test trial looking. Results indicated a significant main effect of trials, $F(2, 30) = 35.90$, $p < .001$, $\eta^2_p = .71$, and planned comparisons revealed significant differences between each trial type, including an increase in looking from posthabituation to test ($p < .001$), reflecting visual recovery to the novel face or voice.

Secondary analyses also assessed whether the gender of the actors affected infants' visual recovery to the novel face or voice. Results of a one-way ANOVA on visual recovery, with gender as a between-subjects factor, indicted no significant effect, $F(1, 14) = 0.13$, $p > .10$, suggesting that gender of the actors did not affect infants' ability to detect a unimodal change in the face or voice.

Results of this study were also compared with those of the 2-month-olds of Experiment 1 who received bimodal habituation and changes (see Table 1). Separate ANOVAs were conducted on baseline, seconds to habituation, mean posthabituation looking, and mean test trial looking to determine whether the patterns of habituation differed across conditions. Results indicated no significant differences as a function of condition (unimodal: Experiment 2 vs. bimodal: Experiment 1; all ps $>$ .10). An ANOVA was also performed on visual recovery as a function of condition to determine whether detection of the unimodal changes in Experiment 2 was greater than detection of the bimodal changes in Experiment 1. Results indicated a significant main effect, with visual recovery to the unimodal changes of Experiment 2 being significantly greater than to the bimodal changes of Experiment 1, $F(1, 35) = 4.18$, $p = .049$. Results of this experiment demonstrate that 2-month-olds are capable of discriminating the faces and the voices used in Experiment 1. Because 2-month-olds detected the unimodal changes in the faces and voices, they either failed to discriminate the faces and voices in the bimodal stimulation of Experiment 1 and/or failed to notice or learn the arbitrary relations between them.

General Discussion

These studies reveal the emergence and development of infants' ability to perceive, learn, and remember the relationship between specific faces and voices of unfamiliar adults speaking. Following a single habituation exposure (3–5 min) to naturalistic films of two adults speaking, infants of 4 and 6 months of age, but not infants of 2 months, showed visual recovery to a change in face–voice relations, indicating they detected the arbitrary intermodal relations between the appearance of each face and the particular sound of each voice. Further, the 4- and 6-month-olds detected the specific face–voice relations regardless of whether the pairs of speakers were male or female.

Ten minutes following the habituation phase, infants of 4 and 6 months were given an intermodal preference test to determine whether they had learned and could remember which face belonged with each voice. Both faces were seen speaking in synchrony with one of the voices that came from a centralized speaker, and thus matching demonstrates acoustically guided visual exploration and intermodal learning of arbitrary face–voice relations. Results indicated that only the 6-month-olds showed intermodal knowledge and memory for face–voice relations. Further, the performance of the 6-month-olds was significantly better than that of the 4-month-olds. Thus, although 4- and 6-month-olds were able to discriminate a change in the arbitrary face–voice relationship (habituation phase), only the 6-month-olds showed evidence of remembering the relations and matching faces and voices on the basis of these relations. Further, additional analyses suggested that the failure of the 4-month-olds to show matching was not due to fatigue (from looking during habituation) nor to interference from viewing mismatching faces and voices during the test trials. Discriminating arbitrary face–voice relations appears to developmentally precede memory for these relations and the ability to use intermodal knowledge of these relations for coordinated audiovisual exploration of events.

The present research also demonstrated that the failure of 2-month-olds to detect unique face–voice relations in the habituation phase was not due to an inability to discriminate among the faces and voices presented. In Experiment 2, 2-month-olds discriminated among all the pairs of faces and voices of Experiment 1. Following habituation to a single face (showing silent speech) or voice (with a static face rather than a synchronously moving face), infants showed visual recovery to a novel face or voice in the unimodal, dynamic stimulation.

Why might 2-month-olds discriminate individual faces and voices in dynamic unimodal stimulation yet not detect a relation between them in dynamic bimodal stimulation? One possibility is that 2-month-olds detect the modality-specific information specifying the appearance of the face and the sound of the voice in bimodal stimulation but fail to relate them to one another (i.e., failure to detect relational information). Another possibility is that it is relatively difficult for 2-month-olds to attend to modality-specific information (such as that specifying the appearance of the face or the sound of the voice) in bimodal stimulation (such as audiovisual speech). This latter possibility is consistent with predictions of the intersensory redundancy hypothesis (Bahrick & Lickliter, 2000, 2002) and recent research supporting this view showing that detection of modality-specific information is facilitated in unimodal rather than bimodal audiovisual events in early infancy (Bahrick, Lickliter, & Flom, in press). For example, 2-month-olds showed unimodal discrimination of faces (during visual speech) but not bimodal discrimination (during audiovisual speech; Bahrick, Lickliter, Vaillant, et al., 2004). Later in development, as attention becomes more flexible, 3-month-olds discriminated among faces even in bimodal audiovisual speech (for a discussion, see Bahrick & Lickliter, 2002; Bahrick, Lickliter, & Flom, 2004). Thus, in Experiment 2 (as in Bahrick et al., 2003; Bahrick, Lickliter, Vaillant, et al., 2004), 2-month-old infants discriminated among the faces and voices in unimodal, nonredundant stimulation but likely could not yet discriminate among them in bimodal stimulation (Experiment 1) where redundant amodal information is more salient and recruits attention. By 4 months of age, however, infants could more easily discriminate among the faces and voices in the bimodal stimulation of Experiment 1.

The present developmental findings are also consistent with an increasing specificity view of perceptual development (Bahrick, 2001, 2004; Gibson, 1969) and a growing body of research demonstrating that detection of arbitrary, modality-specific relations is guided and constrained by the detection of amodal relations uniting audible and visible stimulation (e.g., Bahrick, 1988, 2001; Bahrick & Pickens, 1994; Gogate & Bahrick, 1998; Hernandez-Reif & Bahrick, 2001; Walker-Andrews, 1997). That is, initial detection of global, amodal relations (e.g., voice–lip synchrony, tempo and rhythm of audiovisual speech) appropriately constrains attention and ensures that infants explore patterns of stimulation that belong together. This facilitates further differentiation of unitary multimodal events and in turn promotes differentiation of specific aspects of visual and acoustic stimulation that belong together (e.g., the visual appearance of the face and its relation to the pitch and timbre of the voice). Thus, detection of amodal temporal relations can guide attention to more specific details of multimodal events and lead to the development of intermodal knowledge about unitary audiovisual events.

It is interesting that our results demonstrated that the younger infants (2- and 4-month-olds) spent more time exploring the faces and voices during the habituation phase than did the older infants (6-month-olds). However, this additional time did not appear to facilitate detection of face–voice relations for the 2-month-olds. Exploration may proceed in order of increasing specificity across development as well as within an episode of exploration (Bahrick, 2001, 2004). Thus, when infants encounter bimodal stimulation such as audiovisual speech, they may focus on global information first and for longer periods of time (e.g., voice–lip synchrony, rhythm, tempo, and intensity changes) and then detect modality-specific information. Because younger infants process information more slowly and attention is limited, they may disengage before attending to more specific levels of stimulation such as the appearance of faces and the sound of voices. According to this version of an increasing specificity view, if one could engage the attention of younger infants for a longer period, then detection of face–voice relations would likely emerge in bimodal stimulation.

Overall, this series of experiments reveals a cohesive picture of the development of perception and knowledge about novel and unique faces and voices. Under the present conditions with 3–5-min exposure to the events, by the age of 2 months, infants discriminate among faces and among voices of unfamiliar adults engaged in dynamic speech in unimodal, nonredundant stimulation. By the age of 4 months, infants attend to, perceive, and discriminate among individual faces and voices during naturalistic, bimodal speech. They also detect the unique relationship between the visual appearance of the face and the particular sound of the voice. Thus, intermodal learning about face–voice relations of unfamiliar speakers appears to emerge between 2 and 4 months of age, after infants become skilled at detecting specific faces and voices in multimodal, redundant stimulation. Somewhat later in development, perhaps between 4 and 6 months, infants show intermodal knowledge and memory for face–voice relations and can use their knowledge to guide exploration. They selectively attend to the face that belongs with the voice they are hearing even when other speaking faces are visible. Discrimination of face–voice relations thus precedes and provides a basis for intermodal learning, memory, and exploration of face–voice relations.

Results of the present study also complement and converge with those of prior research on infant perception of face–voice relations. Detection of amodal face–voice relations such as temporal synchrony (e.g., Lewkowicz, 2000), information specifying affect (e.g., Walker-Andrews, 1997), and gender and age of speaker (Bahrick et al., 1998; Walker-Andrews et al., 1991) has been demonstrated for unfamiliar adults by infants between 4 and 7 months of age. However, little is known about infants' perception of arbitrary, modality-specific relations between unique faces and voices. The present study fills this gap in our knowledge. Results converge with and extend those of Brookes et al. (2001) where 3-month-olds showed limited evidence of detecting the face–voice relations in the bimodal stimulation of novel, same-gender adults. The age of 3 to 4 months appears to be a period during which infants can and do attend to the relationship between the specific faces and voices of unfamiliar individuals during naturalistic speech.

The faces and voice of people provide, arguably, the most important and salient source of stimulation for infant attention, perception, learning, and memory. Much perceptual and cognitive development occurs in the context of close face-to-face interaction in the first months of life. The present research adds to growing knowledge about the nature of stimulation infants abstract, learn, and remember from interactions with adults in their environment and how this changes across development. It also highlights the importance of investigating infant attention, learning, and memory for these social events within the context of the multimodal stimulation provided by the social environment.

References

Bahrick, L. E. (1988). Intermodal learning in infancy: Learning on the basis of two kinds of invariant relations in audible and visible events. *Child Development, 59,* 197–209.

Bahrick, L. E. (1992). Infants' perceptual differentiation of amodal and modality-specific audio–visual relations. *Journal of Experimental Child Psychology, 53,* 180–199.

Bahrick, L. E. (1994). The development of infants' sensitivity to arbitrary intermodal relations. *Ecological Psychology, 6,* 111–123.

Bahrick, L. E. (2001). Increasing specificity in perceptual development: Infants' detection of nested levels of multimodal stimulation. *Journal of Experimental Child Psychology, 79,* 253–270.

Bahrick, L. E. (2002). Generalization of learning in three-and-a-half-month-old infants on the basis of amodal relations. *Child Development, 73,* 667–681.

Bahrick, L. E. (2004). The development of perception in a multimodal environment. In G. Bremner & A. Slater (Eds.), *Theories of infant development* (pp. 90–120). Malden, MA: Blackwell.

Bahrick, L. E., Flom, R., & Lickliter, R. (2002). Intersensory redundancy facilitates discrimination of tempo in 3-month-old infants. *Developmental Psychobiology, 41,* 352–363.

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology, 36,* 190–201.

Bahrick, L. E., & Lickliter, R. (2002). Intersensory redundancy guides early cognitive and perceptual development. In R. Kail & H. Reese (Eds.), *Advances in child development and behavior* (Vol. 30, pp. 153–187). New York: Academic Press.

Bahrick, L. E., & Lickliter, R. (2004). Infants' perception of rhythm and tempo in unimodal and multimodal stimulation: A developmental test of the intersensory redundancy hypothesis. *Cognitive, Affective, and Behavioral Neuroscience, 4,* 137–147.

Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science, 13,* 99–102.

Bahrick, L. E., Lickliter, R., & Flom, R. (in press). Up versus down: The role of intersensory redundancy in infants' sensitivity to the orientation of moving objects. *Infancy*.

Bahrick, L. E., Lickliter, R., Shuman, M., Batista, L., & Grandez, C. (2003, April). *Infant discrimination of voices: Predictions from the intersensory redundancy hypothesis*. Poster session presented at the Biennial Meeting of the Society for Research in Child Development, Tampa, FL.

Bahrick, L. E., Lickliter, R., Vaillant, M., Shuman, M., & Castellanos, I. (2004, May). *Infant discrimination of faces in the context of dynamic multimodal events: Predictions from the intersensory redundancy hypothesis*. Poster session presented at the International Conference on Infant Studies, Chicago.

Bahrick, L. E., Moss, L., & Fadil, C. (1996). The development of visual self-recognition in infancy. *Ecological Psychology, 8,* 189–208.

Bahrick, L. E., Netto, D., & Hernandez-Reif, M. (1998). Intermodal perception of adult and child faces and voices by infants. *Child Development, 69,* 1263–1275.

Bahrick, L. E., & Pickens, J. (1994). Amodal relations: The basis for intermodal perception and learning in infancy. In D. J. Lewkowicz & R. Lickliter (Eds.), *The development of intersensory perception: Comparative perspectives* (pp. 205–233). Hillsdale, NJ: Erlbaum.

Barrera, M., & Maurer, D. (1981). Recognition of mother's photographed face by the three-month-old infant. *Child Development, 52,* 714–716.

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.

Birnholtz, J. C., & Benacerraf, B. R. (1983, November). The development of human fetal hearing. *Science, 222,* 516–518.

Brookes, H., Slater, A., Quinn, P. C., Lewkowicz, D. J., Hayes, R., & Brown, E. (2001). Three-month-old infants learn arbitrary auditory–visual pairings between voices and faces. *Infant and Child Development, 10,* 75–82.

Bushnell, I. W. R. (1982). Discrimination of faces by young infants. *Journal of Experimental Psychology, 33,* 298–308.

Carpenter, G. C., Teece, J., Stechler, G., & Friedman, S. (1970). Differential visual behavior to human and humanoid faces in early infancy. *Merrill-Palmer Quarterly, 16,* 91–108.

Cohen, L., & Strauss, M. S. (1979). Concept acquisition in the human infant. *Child Development, 50,* 419–424.

Colombo, J., Mitchell, D. W., Coldren, J. T., & Freeseman, L. J. (1991). Individual differences in infant visual attention: Are short lookers feature processors or faster processors? *Child Development, 62,* 1247–1257.

Cooper, R., & Aslin, R. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61,* 1584–1595.

Cornell, E. (1974). Infants' discrimination of faces following redundant presentations. *Journal of Experimental Child Psychology, 18,* 98–106.

DeCasper, A. J., & Fifer, W. (1980, June). Of human bonding: Newborns prefer their mothers' voices. *Science, 208,* 1174–1176.

DeCasper, A. J., & Prescott, P. A. (1984). Human newborns' perception of male voices: Preference, discrimination, and reinforcing value. *Developmental Psychobiology, 5,* 481–491.

DeCasper, A. J., & Spence, M. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development, 9,* 133–150.

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

Easterbrook, M., Kisilevsky, B., Hains, S., & Muir, D. (1999). Faceness or complexity: Evidence from newborn visual tracking of facelike stimuli. *Infant Behavior and Development, 22,* 17–35.

Fagan, J. F. (1972). Infants' recognition memory for faces. *Journal of Experimental Child Psychology, 14,* 453–476.

Fagan, J. F. (1974). Infant recognition memory: The effect of length of familiarization and type of discrimination task. *Child Development, 45,* 351–356.

Fagan, J. F. (1976). Infants' recognition of invariant features of faces. *Child Development, 47,* 627–638.

Fernald, A. (1984). The perceptual and affective salience of mothers' speech to infants. In L. Feagans, D. Garvey, & R. Golinkoff (Eds.), *The origins and growth of communication* (pp. 5–29). Norwood, NJ: Ablex.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development, 60,* 1497–1510.

Field, T. M., Cohen, D., Garcia, R., & Greenberg, R. (1984). Mother–stranger face discrimination by the newborn infant. *Infant Behavior and Development, 7,* 19–25.

Freeseman, L. J., Colombo, J., & Coldren, J. T. (1993). Individual differences in infant visual attention: Discrimination and generalization of local and global properties at 4 months of age. *Child Development, 64,* 1191–1203.

Frick, J. E., & Colombo, J. (1996). Individual differences in infant visual attention: Recognition of degraded visual forms by four-month-olds. *Child Development, 67,* 188–204.

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

Gibson, E. J., & Pick, A. (2000). *An ecological approach to perceptual learning and development.* New York: Oxford University Press.

Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology, 69,* 1–17.

Hale, S. (1990). A global developmental trend in cognitive processing speed. *Child Development, 61,* 653–663.

Hernandez-Reif, M., & Bahrick, L. E. (2001). The development of visual–tactual perception of objects: Amodal relations provide the basis for learning arbitrary relations. *Infancy, 2,* 51–72.

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

Kleiner, K. A. (1987). Amplitude and phase spectra as indices of infants' pattern preference. *Infant Behavior and Development, 10,* 49–59.

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

Lewkowicz, D. J. (1996). Infants' response to the audible and visible properties of the human face: Role of lexical syntactic content, temporal synchrony, gender, and manner of speech. *Developmental Psychology, 32,* 347–366.

Lewkowicz, D. J. (2000). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin, 126,* 281–308.

Maurer, D., & Barrera, M. (1981). Infants' perception of natural and distorted arrangements of a schematic face. *Child Development, 52,* 196–202.

Maurer, D., & Young, R. E. (1983). Newborns' following of natural and distorted arrangements of facial features. *Infant Behavior and Development, 6,* 127–131.

Miller, C. L. (1983). Developmental changes in male/female voice classification by infants. *Infant Behavior and Development, 6,* 313–330.

Miller, C. L., Younger, B. A., & Morse, P. A. (1982). The categorization of male and female voices in infancy. *Infant Behavior and Development, 5,* 143–159.

Morrongiello, B. A., Lasenby, J., & Lee, N. (2003). Infants' learning, memory, and generalization of learning for bimodal events. *Journal of Experimental Child Psychology, 84,* 1–19.

Nelson, C. A. (2001). The development of neural bases of face recognition. *Infant and Child Development, 10,* 3–18.

Pascalis, O., de Haan, M., Nelson, C. A., & de Schonen, S. (1998). Long-term recognition memory for faces assessed by visual paired comparison in 3- and 6-month-old infants. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 249–260.

Pascalis, O., & de Schonen, S. (1994). Recognition in 3- to 4-day-old human neonates. *NeuroReport, 5,* 1721–1724.

Pascalis, O., de Schonen, S., Morton, J., Deruelle, C., & Fabre-Grenet, M. (1995). Mother's face recognition by neonates: A replication and extension. *Infant Behavior and Development, 18,* 79–85.

Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development, 22,* 237–247.

Querleu, D., Renard, X., Boutteville, C., & Crepin, G. (1989). Hearing by the human fetus? *Seminars in Perinatology, 13,* 409–420.

Rochat, P. (1999). *Early social cognition: Understanding others in the first months of life.* Mahwah, NJ: Erlbaum.

Rose, S. A. (1983). Differential rates of visual information processing in full-term and preterm infants. *Child Development, 54,* 1189–1198.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2002). Processing speed in the first year of life: A longitudinal study of preterm and full-term infants. *Developmental Psychology, 38,* 895–902.

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception and Psychophysics, 59,* 347–357.

Sai, F., & Bushnell, I. W. R. (1988). The perception of faces in different poses by 1-month-olds. *British Journal of Developmental Psychology, 6,* 35–41.

Slater, A., Quinn, P. C., Brown, E., & Hayes, R. (1999). Intermodal perception at birth: Intersensory redundancy guides newborn infants' learning of arbitrary auditory–visual pairings. *Developmental Science, 2,* 333–338.

Soken, N. H., & Pick, A. D. (1992). Intermodal perception of happy and angry expressive behaviors by seven-month-old infants. *Child Development, 63,* 787–795.

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

Stoffregen, T. A., & Bardy, B. G. (2001). On specification and the senses. *Behavioral and Brain Sciences, 24,* 195–261.

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

Walker-Andrews, A. (1994). Taxonomy for intermodal relations. In D. J. Lewkowicz & R. Lickliter (Eds.), *The development of intersensory perception: Comparative perspectives* (pp. 39–56). Hillsdale, NJ: Erlbaum.

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin, 121,* 437–456.

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

Walker-Andrews, A. S., & Grolnick, W. (1983). Discrimination of vocal expressions by young infants. *Infant Behavior and Development, 6,* 491–498.

Walker-Andrews, A. S., & Lennon, E. (1991). Infants' discrimination of vocal expressions: Contributions of auditory and visual information. *Infant Behavior and Development, 14,* 131–142.

Walton, G. E., Bower, N. J., & Bower, T. G. R. (1992). Recognition of familiar faces by newborns. *Infant Behavior and Development, 15,* 265–269.

Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology, 43,* 230–246.

Received May 28, 2004
Revision received September 9, 2004
Accepted November 26, 2004 ■

Instructions to Authors

For Instructions to Authors, please consult the March 2005 issue of the volume or visit www.apa.org/journals/dev and click on the "Instructions to authors" link in the Journal Info box on the right.