Child Development, October 1998, Volume 69, Number 5, Pages 1263–1275

Intermodal Perception of Adult and Child Faces and Voices by Infants

Lorraine E. Bahrick, Dianelys Netto, and Maria Hernandez-Reif

This research investigated the ability of 4- and 7-month-old infants to match unfamiliar, dynamic faces and voices on the basis of age or maturity. In Experiment 1, infants received videotaped trials of an adult and a child of the same gender, side by side, speaking a nursery rhyme in synchrony with one another. The voice to one and then the other face was played in synchrony with the movements of both faces in a random order across 12 trials. On one block of 6 trials a man and a boy were presented, and on the other block a woman and a girl. Results indicated significant matching of the faces and voices at both ages, and the infant's prior experience with children appeared to facilitate matching at 7 months. Further, a visual preference for the children's faces was found. Experiment 2 assessed matching to the same events by 7-month-olds, only with the faces inverted. Results indicated no evidence of matching; however, the visual preference for the children's faces was replicated. Together, the findings suggest that infants are able to detect invariant intermodal relations specifying the age or maturity of a person's face and voice. This matching was most likely based on information that was degraded by inverting the faces, including invariant relations between the sound of the voice and configurational aspects of the face, or between temporal aspects of the voice and the relative motion of facial features.

INTRODUCTION

Infants are remarkably skilled perceivers of invariant relations in audible and visible stimulation. They are able to abstract temporal relations such as voice-lip synchrony (Dodd, 1979), synchrony between visual and acoustic impacts (Bahrick, 1983, 1988), and between sounds and changes in the direction of motion (Lewkowicz, 1992; Spelke, Born, & Chu, 1983) as well as the common rhythmic structure, tempo, and duration of audible and visible events (Allen, Walker, Symonds, & Marcell, 1977; Lewkowicz, 1986; Mendelson & Ferland, 1982; Pickens & Bahrick, 1995; Spelke, 1979; see Bahrick & Pickens, 1994, for a review). These temporal relations are amodal; the same temporal information can be both seen and heard. It is redundant across the two sense modalities. Infants are adept at detecting amodal invariant relations, and detection of these relations may developmentally precede detection of arbitrary intermodal relations, providing a natural buffer against learning incongruent object-sound relations (Bahrick, 1992, 1994; Bahrick & Pickens, 1994).

Some audible and visible properties are specified by more complex classes of natural invariant relations. For example, Walker-Andrews, Bahrick, Raglioni, and Diaz (1991) found that young infants were sensitive to bimodal information specifying the gender of adults. The invariant information was described as a natural and "typical" multimodal relation. These relations occur in nature with regularity, but are neither exclusively amodal nor arbitrary (see also Walker-Andrews, 1994). They may be comprised of sets of overlapping invariants, some amodal and some more arbitrary. For example, male faces are larger and their voices are deeper than those of females, but there is some overlap between the two genders with respect to these attributes. Further, we can often identify the gender of an individual on the basis of other conventional, arbitrary relations such as hair style and length, clothing, jewelry, and makeup.

The present research investigates the development of infants' perception of another complex class of "typical" multimodal relations, that defining the age of an individual. We asked whether infants could match the faces and voices of adults versus children on the basis of invariant information specifying the age or maturity of the person. The dimension of age provides overlapping sets of invariant relations, both amodal and arbitrary. Children typically have smaller, rounder heads and voices of a higher pitch than adults. But, similar to the case of gender, there is a much more complex set of relations that specify "babyishness" (see, Alley, 1981, 1988; Enlow, 1982). Compared to adults, children may have a more round head shape; more filled-out cheeks; smaller, more rounded features; faces with jaws and mouths that appear smaller and eyes that appear larger; skin of a softer texture; movements that are more labile, less well controlled, and that show a greater intensity range; and their voices have a higher pitch and are less well controlled, more labile, and have a greater pitch and amplitude range. Infants have a great deal of experience with faces and voices early on and may thus show early perceptual learning for this complex "typical" relation. Further, children and adults may afford different interactions for infants, and thus this distinction may be quite meaningful to them with respect to action consequences, especially for infants with older siblings.

Prior research has demonstrated that infants are quite capable of discriminating among voices. Even neonates are able to discriminate between their mothers' voices and those of female strangers (DeCasper & Fifer, 1980) and between their fathers' voices and those of male strangers, or between two unfamiliar male voices (DeCasper & Prescott, 1984). Infants also prefer to listen to infant-directed speech over adult-directed speech, which differ from each other along dimensions such as pitch, tempo, and intonation pattern (e.g., Cooper & Aslin, 1990; Fernald, 1985; Werker & McLeod, 1989). Further, they differentiate and classify voices on the basis of gender (Miller, 1983; Miller, Younger, & Morse, 1982).

Research in the area of face perception (see Nelson & Ludemann, 1989, for a review) has demonstrated the ability to discriminate and categorize static representations of faces in early infancy for mother versus stranger (e.g., Barrera & Maurer, 1981; Bernard & Ramey, 1977; Bushnell, 1982; Caron, Caron, Caldwell, & Weiss, 1973) and among unfamiliar faces as well (Cohen & Strauss, 1979; Cornell, 1974; Fagan, 1972, 1976). Even newborns preferred a still image of the mother's face over that of a female stranger (Walton, Bower, & Bower, 1992). When presented with dynamic or live faces, neonates and infants during the first month are able to discriminate the mother from a female stranger (e.g., Carpenter, Teece, Stechler, & Friedman, 1970; Field, Cohen, Garcia, & Greenberg, 1984; Sai & Bushnell, 1988). By 4–7 months, babies show discrimination among different facial expressions (Biringen, 1987; Caron, Caron, & MacLean, 1988; Kestenbaum & Nelson, 1990; Nelson & Horowitz, 1983), and perhaps even at birth (Field et al., 1983). Infants also demonstrate a preference for "attractive" faces (Langlois, Ritter, Roggman, & Vaughn, 1991; Samuels, Butterworth, Roberts, Graupner, & Hole, 1994). Further, infants have been shown to be quite sensitive to dynamic information specifying a face in the absence of featural information (Stucki, Kaufmann-Hayoz, & Kaufmann, 1987). Thus, infants show a diverse and well developed set of skills for perceiving faces and understanding the information conveyed by them in the first half year of life.

Intermodal perception of face-voice relations has received less attention in the literature. Nine- and 12-month-olds showed preferential looking to black-and-white photos of women's faces when women's voices as compared with men's voices were played (Poulin-Dubois, Serbin, Kenyon, & Derbyshire, 1994). Intermodal perception appears to be more robust, to manifest earlier, and to be based on a variety of attributes when dynamic displays are presented (see Lewkowicz, 1996; Walker-Andrews, 1997). By 2 months, infants relate dynamic faces and voices on the basis of voice-lip synchrony (Dodd, 1979), at 4 months on the basis of the spectral information in vowel sounds (Kuhl & Meltzoff, 1984), and by 5–7 months on the basis of the common affective expression (Walker, 1982; Walker-Andrews, 1988, 1997). Matching visual and acoustic affect can also be accomplished on the basis of motion information without the help of featural information, as evidenced by infants' matching point-light displays of faces with appropriate vocal expressions (Soken & Pick, 1992). Finally, as discussed earlier, infants relate faces and voices on the basis of gender (Walker-Andrews et al., 1991). When films of a male and a female adult were played side by side reciting a nursery rhyme in synchrony with one another, 4- and 6-month-old infants were able to match the faces with the appropriate voices on the basis of gender of speaker. They looked significantly more to the male face when the male voice was played and to the female face when the female voice was played. These relations were detected when temporal parameters such as voice-lip synchrony were controlled.

There is little research addressing infants' perception of adults versus children, even though this is a significant dimension from the point of view of the infant. Adults may afford caregiving and comfort, whereas young children often do not; they may more often afford play and social interaction. Only one study assessing intermodal perception of face-voice relations on the basis of age has appeared. Lasky, Klein, and Martinez (1974) assessed whether 5- to 6-month-old Guatemalan infants were able to relate static faces and voices of adults and children. Results indicated intermodal matching of faces and voices when the photographs of a woman and a 7-year-old boy were presented, but not when face-voice pairs of the same gender were presented. This, however, may have been due to a failure to discriminate among the same-sex photos. With respect to infant perception of strangers, Greenberg, Hillman, and Grice (1973) reported that 8- and 12-month-old infants showed significantly more stranger anxiety toward adults than toward children aged 4–5. Other studies reported similar findings (Bigelow, MacLean, Wood, & Smith, 1990; Brooks & Lewis, 1976; Lewis & Brooks, 1974). Infants' differential response to children versus adults could be based on any number of variables, including motion of the face and/or body, simple featural differences between the faces of adults and children, or more complex invariant relations that differentiate the facial configurations of adults and children.

The present research investigates the development of infants' sensitivity to intermodal information specifying age uniting the dynamic faces and voices of unfamiliar adults and children. Perception was assessed for same-sex pairs to rule out gender as a basis of matching. We used an intermodal preference procedure similar to that of our prior study (Walker-Andrews et al., 1991, Exp. 2), where two faces were presented side by side along with a voice appropriate to one, but synchronized with the motions of both. This procedure can provide us with a window into the infants' discrimination of dynamic faces of adults versus children, discrimination of voices, and their ability to relate unfamiliar faces and voices when voice-lip synchrony and tempo of action are controlled. Successful matching may imply more than discrimination and matching of simple features because age, like gender, is defined by a complex class of overlapping properties. Because infants of 4 months were found to show attenuated evidence of face-voice matching on the basis of gender in this procedure, this was the younger age chosen for testing in the present study. Seven months served as the older age, because by this time infants typically show categorization on the basis of a number of attributes such as gender, identity of an individual across different poses, speech sounds, rhythm and tempo of auditory sequences, and so on (e.g., Cohen & Strauss, 1979; Fagan, 1976; Grieser & Kuhl, 1989; Pickens & Bahrick, 1997; Trehub & Thorpe, 1989), and matching of faces and voices on the basis of gender, possibly an easier task, was found to be present by 6 months in our prior study.

EXPERIMENT 1

Four- and 7-month-old infants were presented with the faces of an adult and child of the same gender side by side, speaking in synchrony. On some trials, the voice belonging to the child was presented, and on some trials, the voice of the adult was presented in synchrony with the motions of both faces. The synchronous presentation controlled for potential confounds such as matching or discrimination on the basis of tempo of action or voice-lip synchrony. In addition, an effort was made to equate the adult and child actors for facial and vocal affect by having all of the actors match their speech and facial expressions to a videotaped model. This allowed us to eliminate these simple differences as a basis of intermodal matching. Further, matching on the basis of gender was also eliminated because adult-child pairs were of the same gender.

Method

Participants. Sixteen 4-month-old infants (10 males and six females) whose mean age was 126 days (SD = 6.4) and 16 7-month-olds (11 males and five females) whose mean age was 222 days (SD = 5.5) participated. All were healthy, normal, term infants, weighing at least 5 pounds at birth, with Apgar scores of 7 or higher. The data of 12 additional infants were rejected from the study (N = 8 at 4 months, and N = 4 at 7 months). At 4 months, two were rejected for excessive fussiness and six for failure to meet the attention criterion (see Procedure section for details). At 7 months, one participant's data were rejected for equipment failure and three for excessive fussiness.

Stimulus materials. Color video films of a boy and girl reciting a nursery rhyme were made and modeled after those of the adult male and female actors used by us in the Walker-Andrews et al. (1991) study on gender perception. The final events consisted of two 9-year-old children and two adults, aged 25–30 years. Both males had fairly short, dark hair, and both females had shoulder-length, dark hair. All actors were filmed (with a Panasonic WV-3170 camera) speaking a nursery rhyme, "This Old Man," for a 2 min period, and the films depicted a view of their faces and shoulders. The natural variations in head sizes were preserved, such that on the 19 inch video monitors, the head sizes from the bottom of the chin to the top of the crown measured 20.5, 19.5, 17.5, and 17.5 cm for the adult male, female, and child male and female, respectively. Synchrony between the child and adult speech was accomplished by having the children (after practice) recite the nursery rhyme in synchrony with a videotaped model; the same videotaped model was used by the adult actors in the Walker-Andrews et al. (1991) study. Children were instructed to synchronize their lip motions, and to incorporate the facial expressions and affect of the model, specifically, the timing of the smiles and eyebrow movements at the beginning of each verse. This was done to roughly equate for affect and overall amount of motion across actors. Finally, to ensure that the degree of face-voice synchrony was similar across conditions, each actor dubbed his or her speaking voice onto the video image of his or her speaking face. Thus, when adult and child films were played side by side, the lip movements of all actors were in very good synchrony with both faces, but none was in perfect synchrony. A number of children were pretested for these skills before two were selected.

Apparatus. Infants sat in a standard infant seat facing two color video monitors (Panasonic BT-D1920Y) approximately 50 cm away. A column of colored lights and a mechanical toy dog were located between the monitors to attract the infants' attention between trials. Two apertures were cut into the black posterboard surrounding the monitors and were used by two observers to watch the infants' visual fixations to the displays. Observers recorded fixations by depressing one of a set of two buttons while the infant fixated the right or left displays. The fixations were recorded on an eight channel Rustrak strip-chart recorder.

The video displays were presented using two video decks (Panasonic AG-6300) that were connected to an edit controller (Panasonic NV-A500). The edit controller enabled precise synchronization (to the nearest .02 s) of the output from the two video decks so that the lip movements of the adult and child speakers could be kept aligned throughout the study. The stimulus voices were projected through a speaker centered midway between the two monitors and just below them at approximately 55 db.

Procedure. Prior to participating, parents completed a questionnaire regarding the number of siblings in the family and the amount of interaction the infant typically had with children. This served as a basis for estimating the amount of experience the participants had with young children.

Infants were presented with 12 20 s trials comprised of two blocks of six trials each. Each trial depicted an adult and a child speaking in synchrony side by side, along with the synchronized voice matching one of them. In each block, the child's soundtrack was played on three trials and the adult's soundtrack on three trials in a semirandom order (such that a given voice occurred no more than twice in succession). The same semirandom order was repeated for the second block of trials. Gender of the adult-child pair was manipulated as a within-subjects factor. For one block of trials, infants received the male adult-child pair, whereas for the other block they received the female adult-child pair. The order of presentation of the two blocks was counterbalanced. The lateral positions of the adult and child films were also counterbalanced within participants in each age group such that half saw adult right/child left for the first block and adult left/child right for the second block, or vice versa. Intertrial intervals averaged 4 s, and interblock intervals were approximately 90 s.

Two looking criteria were set for the data of participants to be included in the experiment. They were required to complete five out of the six trials in each block. No data were rejected for failure to meet this criterion. We also felt it was important that participants notice that there were in fact two video displays on the first trial of each block. Thus, an attention criterion was set requiring that infants look at each stimulus at least 1 s on the first trial of each block. The data of six of the 4-month-olds were eliminated for failure to look at the second film.

Trained observers who were unaware of the lateral positions of the video displays monitored the infants' looking times. The second observer recorded looking for 12 of the 32 participants ($N = 6$ at each age). Interobserver reliability, expressed as a Pearson product-moment correlation between the looking proportions of the primary and secondary observer, averaged .96.

Results

The mean number of seconds that infants at each age spent looking at the visual displays during Block 1 (trials 1–6) and Block 2 (trials 7–12) are depicted in Table 1. A two-way analysis of variance with age as a between-subjects factor and block as a within-subjects factor indicated a significant main effect of block, $F(1, 30) = 18.16$, $p < .001$, and no main effect of age or age × block interaction ($ps > .1$). Thus, infants at both ages showed a decrease in their overall looking time from Block 1 to Block 2.

The primary results were expressed in terms of the proportion of total looking time (PTLT) infants looked to the sound-matched video display. Proportions were derived for each trial separately and then averaged to obtain the mean proportion for Block 1 and Block 2. A mean PTLT was also derived by averaging across the two blocks for each infant. Difference scores, the main dependent variable, were calculated for each trial block by subtracting the PTLT to a given face when the voice was mismatched from the PTLT to that face when the voice was matched for each infant. Thus, positive difference scores indicate greater looking to the voice-matched face, whereas negative difference scores indicate greater looking to the voice-mismatched face. These results are displayed in Table 2. In the primary analyses, single-sample t tests were used to determine whether infants showed significant matching of the soundtrack to one of two films by comparing the mean difference

Table 1 Mean Seconds of Looking Time and Standard Deviations to the Matching versus Mismatching Displays during Each Trial Block for 4- and 7-Month-Old Infants

	Block 1			Block 2			Blocks 1 + 2
Age	Match	Mismatch	Total	Match	Mismatch	Total	Total
4 months:							
M	48.0	51.4	99.4	44.8	39.0	83.8	183.2
SD	11.2	9.0	12.8	13.2	10.9	21.5	28.6
7 months:							
M	51.3	38.3	89.6	41.1	38.3	79.4	169.0
SD	12.8	9.2	15.3	10.5	9.2	16.0	28.8

score against chance (0 difference). In secondary analyses, we compare the degree of matching across groups with analyses of variance or two-sample t tests. All reported significance levels are two-tailed and .05 or less.

To address the main research question, whether infants showed significant matching of faces and voices at either age, the difference scores were evaluated separately for each age and block using single-sample t tests against the chance value of 0, with variance pooled across age (given that homogeneity of variance requirements were met according to Cochran C tests, all $ps > .1$). Results indicated that the 7-month-olds showed significant matching of voices and faces during Blocks 1 and 2 combined, $t(15) = 3.60$, $p < .01$, as well as during Block 1 taken alone, $t(15) = 3.12$, $p < .01$, but no evidence of matching was found during Block 2, $t(15) = 1.31$, $p > .1$. In

Table 2 Mean Proportions of Total Looking Time (PTLT), Standard Deviations (*SD*), and Difference Scores (Diff.) for the Adult and Child Films When Each Was Sound-Matched versus Sound-Mismatched

Film:	Child		Adult		Average of Child/Adult		
Voice:	Match Child	Mismatch Adult	Match Adult	Mismatch Child	Match	Mismatch	Diff.
7 months:							
Block 1:							
PTLT	.68	.55	.45	.32	.57	.44	.13**
SD	.21	.21	.21	.21	.08	.08	.17
Block 2:							
PTLT	.63	.59	.41	.37	.52	.48	.04
SD	.22	.17	.17	.22	.06	.06	.11
Blocks 1–2:							
PTLT	.66	.57	.43	.34	.55	.46	.09**
SD	.11	.08	.08	.11	.06	.06	.12
4 months:							
Block 1:							
PTLT	.52	.56	.44	.48	.48	.52	.04
SD	.22	.22	.22	.22	.08	.08	.16
Block 2:							
PTLT	.57	.50	.50	.43	.54	.47	.07*
SD	.24	.24	.25	.25	.06	.06	.13
Blocks 1–2:							
PTLT	.55	.54	.46	.45	.51	.50	.01
SD	.10	.12	.12	.10	.04	.04	.08

Note: The difference score is the proportion of total looking time to a given film when it was sound-matched minus the proportion of total looking time to that film when it was sound-mismatched. The difference score is identical regardless of whether it is calculated on the basis of looking to the child films (columns 1, 2) or looking to the adult films (columns 3, 4), because the scores are reciprocally related and the proportion looking to the child film on a given trial defines the proportion looking to the adult film.

* $p < .05$; ** $p < .01$.

contrast, 4-month-olds showed no overall matching for Blocks 1 and 2 combined, $t(15) = .36$, $p > .1$, or for Block 1 alone, $t(15) = .96$, $p > .1$. However, they did show significant matching on Block 2 alone, $t(15) = 2.29$, $p < .05$. Thus, at both ages, infants were able to match faces and voices of children and adults. Further, results of binomial tests supported these findings. They indicated that a significant proportion of the 7-month-olds showed matching (PTLTs > .50) on Block 1 (12 of 15, one showed no preference, $p = .014$), and a significant proportion of the 4-month-olds showed matching on Block 2 (11 of 15, one showed no preference, $p = .042$).

Secondary analyses were conducted to examine differences across age and trial block. A two-way analysis of variance was performed on the difference scores with age as the between-subjects factor and trial block as the within-subjects factor. Results indicated a significant age × trial block interaction, $F(1, 30) = 7.89$, $p = .009$, no significant main effect of trial block, $F(1, 30) = .13$, $p > .1$, and a marginally significant main effect of age, $F(1, 30) = 3.67$, $p = .065$. Thus, the pattern of matching across trial blocks differed as a function of age. As can be seen from the difference scores in Table 2, the 4-month-olds showed an increase in looking to the matching film from Block 1 to Block 2, whereas the 7-month-olds showed a decrease from Block 1 to Block 2. This may reflect the fact that older infants catch on quickly and become less interested in following the matching film by the last block, whereas the younger infants take more time to detect the audiovisual relations. This pattern parallels that found in prior studies (Bahrick, Moss, & Fadil, 1996; Walker-Andrews et al., 1991) where the older infants showed effects during the first block of trials and became less interested in matching by the second block, whereas younger infants required more exposure time for the significant effects to emerge.

Further analyses assessed whether infants showed any preference for the films of the children or the adults, independent of sound manipulation. Single-sample t tests on the PTLTs to a film regardless of whether it was sound-matched or sound-mismatched revealed a significant preference for the child films at 7 months, $M = .61$, $t(15) = 5.87$, $p = .001$. At 4 months the means were in the same direction, but the preference did not reach significance, $M = .545$, $t(15) = 1.81$, $p = .09$. The difference between the preference at 7 and 4 months was significant, $t(14) = 2.14$, $p = .041$, indicating that 7-month-olds had a significantly greater tendency to watch the child faces than did the 4-month-olds.

Additional secondary analyses for side preferences compared the PTLTs to the screen on the right side, regardless of sound condition, against the chance value of .5. Results indicated no significant preference at 4 or 7 months, $t(15) = .46$, $p > .1$; $t(15) = .49$, $p > .1$, respectively. In addition, preferences for the female versus the male faces were assessed by comparing the total seconds of looking to the female versus the male pair across the two blocks of trials. Results revealed no significant preferences at 4 or 7 months, $t(15) = .42$, $p > .1$; $t(15) = .09$, $p > .1$, respectively.

In analyses assessing the effects of the infants' exposure to children on their matching performance, infants of each age were classified as "high" or "low" in experience according to results of a questionnaire. Infants whose parents reported that they interacted "often" or "most of their waking time" with a child (typically a sibling) were classified as "high" in experience ($N = 9$ at 7 months and $N = 8$ at 4 months). Infants who did not have siblings and whose parents reported that they "never" or "occasionally" interacted with a child were classified as "low" in experience ($N = 7$ at 7 months and $N = 8$ at 4 months). T tests on the difference scores (against chance) were conducted to determine which groups showed evidence of matching. They revealed that the experienced infants at 7 months showed significant matching on Blocks 1 and 2 averaged, $M = .14$, $t(8) = 3.35$, $p = .01$, and on Block 1 alone, $M = .18$, $t(8) = 2.85$, $p = .02$, whereas the inexperienced infants did not, $M = .04$, $t(6) = 1.21$, $p > .1$; $M = .07$, $t(6) = 1.27$, $p > .1$, respectively. Thus, although infants with both high and low experience showed effects in the direction of matching, the high-experience group was carrying the matching effect. In contrast, at 4 months, there appears to be no evidence of experience affecting intermodal matching. In Block 2 (where matching had been evident), neither the high- nor the low-experience groups showed significant matching alone, and their means appear comparable, $M = .08$, $t(7) = 1.45$, $p > .1$; $M = .065$, $t(7) = 1.47$, $p > .1$, respectively. Finally, an analysis of variance on difference scores with age and experience as between-subjects factors and trial block as a within-subjects factor indicated no main effects or interactions (all $ps > .1$). Thus, overall, participants with high versus low experience did not differ from one another in their matching.

The effects of experience on the visual preference for the child films was also assessed. A three-way analysis of variance with age and experience as between-subjects factors and trial block as a within-subjects factor was conducted on the proportion of total looking time to the child films regardless of the sound condition (overall $M = .61$, $SD = .08$ at 7 months; $M = .545$, $SD = .10$, at 4 months). Results revealed a significant main effect of age, $F(1, 28) = 5.12$, $p = .032$, and no main effects of experience, $F(1, 28) = .67$, $p > .1$, trial block, $F(1, 28) = .15$, $p > .1$, or interaction between age and experience, $F(1, 28) = .64$, $p > .1$. The older infants showed a greater child preference than the younger ones, but experience as rated by the parent did not influence this preference.

Discussion

In sum, infants demonstrated the ability to match the faces and voices of children and adults of the same gender, even though voice-lip synchrony was controlled. At 7 months, infants showed this ability overall and in the first block of trials. At 4 months, this ability emerged by the second block of trials. These results parallel those found for matching on the basis of gender (Walker-Andrews et al., 1991), where the 6-month-olds showed matching from the start, whereas the 4-month-olds showed eventual matching. Apparently it takes the younger infants more time to abstract the information that serves as the basis for matching than it does the older infants. Further, the infant's experience with children apparently facilitates abstraction of this information at 7 months. The effect of experience for younger infants was not evident in the present study. To draw more firm conclusions regarding this effect, one would need to test a larger sample of infants and manipulate experience a priori.

What might be the basis for matching the faces and voices? The dynamic presentation of adults and children speaking differed in a number of ways. Some potential explanations are simple: They tap general intermodal abilities and do not require that the visual display be perceived as a facial configuration as such. Other explanations are more theoretically interesting and require that infants perceive the display as a coherent configuration with features that are positioned relative to one another and move relative to one another in particular ways. They require matching on the basis of configurational aspects of the stimulation.

The more simple explanations include matching on the basis of synchrony, pitch-size relations, and/or intensity relations. (1) Minor voice-lip synchrony differences may have been detectable between children and adults. That is, the adult actors may have been better at synchronizing their voices with their videotaped faces than the children, and this could have served as a basis for matching. Further, despite our procedural precautions for equating for minute synchrony differences, the child voice may have been better synchronized with the child face than with the accompanying adult face, and vice versa for the adult voice. Although unlikely as a basis for intermodal matching, it cannot be ruled out in this study. (2) Infants may have matched on the basis of size-pitch relations. Large objects typically make deeper sounds than small objects. Adults typically have larger heads and deeper voices than children. The natural head size variations were preserved in our video representations of the actors. (3) Infants may also have matched on the basis of intensity relations. Children may be more energetic and show more head and face movement as well as more vocal movement and variability than adults. Matching on the basis of hypotheses 1, 2, and 3 would be possible for any kind of audiovisual display. Matching need not be specific to faces and voices, and this information would thus be available in inverted faces.

More theoretically interesting bases for matching the faces and voices are also possible. Matching may have been based on more complex, global face-voice relations that require attention to the spatial configuration and/or temporal changes in the configuration of features relative to one another. Such changes provide information for emotion, gender, identity, and so on and would be much more difficult to detect in inverted faces where the relative positioning of features is disrupted (e.g., Diamond & Carey, 1986). For example, (4) infants may have matched the facial configuration that looked "babyish" with the one that sounded "babyish." Children's faces and features are smaller and rounder, with a configuration that is more compressed and rounded than that of adults. Their voices are also of a higher pitch and have resonance qualities different from those of adults. Matching may thus have been based on invariant spatial aspects of the facial configuration with respect to the sound quality of the voice. Finally, (5) infants may have matched a facial configuration that moved in a "babyish" way with vocal patterns that sounded "babyish." Children may display facial/body motion and vocal patterns that differ from those of adults and that look and sound young. Children's face and body movements may be more labile and less well controlled, with facial expressions conveying youthful affect. Similarly, their voices and intonation patterns may be less well controlled and show greater lability, temporal variability, and more youthful affect than those of adults. Some intonation contours sound more babyish than others, with specific patterns of variability in pitch and facial expressions. The quality of this visual and vocal variability may be invariant and specify age. Matching may thus have been based on the temporal and spatial patterning of relative motion of features with respect to the temporal aspects of speech. Hypotheses 4 and 5 require perceiving the relative positioning of the features of the face and the facial configuration as a coherent and meaningful pattern, whereas hypotheses 1, 2, and 3 do not necessarily require perceiving the face as a whole. Experiment 2 was designed as a first step in evaluating the five hypotheses.

EXPERIMENT 2

This study was designed to examine whether the detection of bimodal face-voice relations was based exclusively on the intermodal relations, voice-lip synchrony differences, pitch-headsize relations, and intensity relations, or whether it was based on more complex, global relations specific to facial configurations. By inverting the video images, one can disrupt perception of configurational aspects of the face while preserving all of the physical information, making perception of emotional expressions and person identity difficult (Diamond & Carey, 1986; Fagan, 1979; Kestenbaum & Nelson, 1990; Oster, 1981; Stucki et al., 1987; Walker, 1982). Thus, detection of intermodal relations that do not necessarily rely on the perception of relations among facial features should still be possible (hypotheses 1, 2, and 3). Given that 7-month-olds showed the strongest matching effects in Experiment 1, this follow-up study was undertaken with 7-month-olds only. Participants were presented with video images of an adult and child, side by side, according to the prior procedures, except that the moving faces were inverted. It was expected that if matching was primarily based on (1) voice-lip synchrony and/or (2) pitch-size relations, and/or (3) intensity relations, infants would continue to show matching under the inverted condition. However, if it was primarily based on matching (4) a babyish versus adult sound and appearance, or (5) a babyish versus adult motion pattern and vocal pattern, infants would no longer show matching, suggesting that for these characteristics, an upright face was necessary for matching.

Method

Participants. Sixteen normal, healthy 7-month-old infants (nine males and seven females) whose mean age was 223 days ($SD = 11.9$) participated. The same criteria used in Experiment 1 for participant selection and inclusion of data were used here. The data of only one additional infant were excluded from the study due to experimenter error. The trained observers of the prior study also participated in this study.

Procedure. The stimulus events, apparatus, and procedures were identical to those of the prior study except that the video displays were presented upside down. This was accomplished by inverting the video monitors.

Results and Discussion

The total number of seconds spent looking at the displays for Blocks 1 and 2 are depicted in Table 3. A matched sample t test was conducted to determine whether looking decreased from Block 1 to Block 2. No significant difference was found, $t(15) = 1.01, p > .1$. Further, a two-sample t test was conducted to compare the total number of seconds infants spent looking at the inverted displays in this study with the total number of seconds spent looking at the upright displays in Experiment 1. Results revealed no significant difference, $t(15) = 1.46, p > .1$, indicating that infants' interest level in the two types of displays did not differ. The PTLTs and difference scores reflecting matching were calculated as before and are displayed in Table 4.

The difference scores for Block 1, Block 2, and Blocks 1 and 2 averaged were tested with single-sample t tests against chance to assess evidence of intermodal matching. No evidence of matching was found for any of the measures (all $ps > .1$). Thus, in the present study, when inverted images were presented, 7-month-olds showed no evidence of matching the faces and voices of children and adults as they had in Experiment 1 when the images were upright. The information that served as the basis for matching was apparently degraded by inverting the faces. Results of a binomial test were also consistent. Only eight of the 16 infants showed matching (PTLTs > .5), and three showed no preference. This was not significantly different from chance ($p > .1$).

The difference scores for Blocks 1 and 2 averaged were also compared across Experiments 1 and 2 using a t test to evaluate whether the change in orientation of the face (from upright to inverted) caused a decrement in matching. Inverting the faces led to a marginally significant decrease in matching the faces and voices, $t(30) = 1.87, p = .08$, from a mean of .09 ($SD = .12$) with upright faces, to a mean of .02 ($SD = .12$) with inverted faces.

Further, the difference scores were analyzed to determine whether intermodal matching was influenced by the infant's experience with children because there was some evidence that experience influenced the matching of 7-month-olds in Experiment 1. Single-sample t tests on the mean difference scores across Blocks 1 and 2 for infants with high ver-

Table 3 Mean Seconds of Looking Time and Standard Deviations to the Matching versus Mismatching Inverted Displays during Each Trial Block

	Block 1			Block 2			Blocks 1 + 2
	Match	Mismatch	Total	Match	Mismatch	Total	Total
M	39.9	38.0	77.9	34.6	34.9	69.5	147.4
SD	15.6	16.4	27.0	14.8	14.3	27.8	51.5

sus low experience with children ($M = .03, M = .005$, respectively) failed to reveal a significant matching effect for the infants with high or low experience, $t(7) = .7, p > .1$; $t(7) = .13, p > .1$, respectively. Thus even infants with high experience did not show intermodal matching with inverted faces.

Further analyses were conducted to determine whether infants showed any preferences for the films of the children over the adults, as in the prior study. Surprisingly, results again indicated a significant preference across the two trial blocks for the child films regardless of the sound condition, $M = .59$, $t(15) = 3.62, p = .003$. Thus, although the faces were inverted and infants were not able to match the voices with the faces, they still preferred to look at the child faces more than the adult faces. Apparently, the information that served as the basis for preferring and discriminating the child from the adult faces was not degraded by inverting the images.

To determine whether the above preference for the child faces was influenced by the infant's experience with children, additional analyses were performed. *T* tests on the PTLT to the child films regardless of sound condition indicated that both infants with high and low experience tended to prefer the child faces, $M = .59, SD = .08, t(7) = 3.17, p = .016$; $M = .59, SD = .13, t(7) = 2.10, p = .074$, respectively. Thus the child preference was not affected by experience with children.

Secondary analyses assessed effects of side preferences independent of sound condition and preferences for the female over the male faces as before. Results indicated no preference for the right side, $M = .48, t(15) = .47, p > .1$, or preference based on gender, $t(15) = .55, p > .1$.

GENERAL DISCUSSION

Experiment 1 revealed that 7- and 4-month-old infants are capable of matching the faces and voices of adults and children. Seven-month-olds showed matching in the first block of trials and 4-month-olds showed matching by the second block of trials. Infants at both ages were able to detect the intermodal

Table 4 Mean Proportions of Total Looking Time (PTLT), Standard Deviations (*SD*), and Difference Scores (Diff.) for Inverted Adult and Child Films When Each Was Sound-Matched versus Sound-Mismatched

	Child		Adult		Average of Child/Adult		
	Match Child	Mismatch Adult	Match Adult	Mismatch Child	Match	Mismatch	Diff.
Block 1:							
PTLT	.64	.61	.39	.36	.52	.49	.03
SD	.17	.20	.20	.17	.10	.10	
Block 2:							
PTLT	.54	.54	.46	.46	.50	.50	.00
SD	.16	.22	.22	.16	.07	.07	.15
Blocks 1–2:							
PTLT	.60	.58	.42	.40	.51	.49	.02
SD	.11	.13	.13	.11	.06	.06	.12

Note: The difference score (Diff.) is the proportion of total looking time to a given film when it was sound-matched minus the proportion of total looking time to that film when it was sound-mismatched. The difference score is identical regardless of whether it is calculated on the basis of looking to the child films (columns 1, 2) or looking to the adult films (columns 3, 4), because the scores are reciprocally related and the proportion looking to the child film on a given trial defines the proportion looking to the adult film.

relations uniting the moving faces and voices and discriminate the faces and the voices of adults from those of children. These findings are similar to those of our prior study assessing infants' bimodal perception of gender (Walker-Andrews et al., 1991); intermodal matching of faces and voices on the basis of gender was also found in infants as young as 4 months by the second block of trials and in 6-month-olds from the first block.

Experiment 2 was conducted to narrow down the number of hypotheses regarding the basis for the intermodal matching of adult and child faces and voices. By inverting the images, we disrupted perception of the facial configurations while preserving the physical information in the displays. Consequently, 7-month-olds no longer showed matching of the faces and voices. Thus, the matching observed in Experiment 1 was most likely based on perceiving the configurational information in the face and/or the relative movement of features in relation to the vocal information. In contrast, the earlier matching was probably not based on simple differences in voice-lip synchrony between the sound-appropriate and sound-inappropriate displays because voice-lip synchrony was preserved in the inverted displays. Nor is it likely that matching was based on detection of pitch-size relations. Larger objects often make sounds that are deeper than those produced by smaller objects. However, this relation could have been detected just as easily with inverted images as upright images. Further, matching was not likely based on detection of intensity relations between the visual and vocal information. Although the faces and voices of children may show more motion or overall variability than those of adults, this information was also preserved in inverted displays. Rather, the basis for matching is apparently more complex and may depend on perceiving the face as a whole, that is, as a configuration and/or the relative motion of features. For example, infants could have detected invariant aspects of the relation between the appearance of the face and the sound or resonance qualities of the voice. Children sound and look more babyish or immature than adults.

Second, infants could have detected invariant relations between the relative motion of the head and facial features in relation to the temporal or prosodic aspects of the voice. Children move and speak in ways that appear more "babyish" than adults. They show distinctive patterns of relative motion of facial features and distinctive vocal patterns that convey youth. They also show visual and vocal affect that convey youth. Research has indicated that infants are quite capable of perceiving relative motion in the human form (Bertenthal, Proffitt, Spetner, & Thomas, 1985; Fox & McDaniel, 1982; Stucki et al., 1987). Motion carries information about properties of faces such as gender, emotional expression, identity, and so on (e.g., Bassili, 1978), and infants are able to discriminate among point light displays on the basis of these attributes (Soken & Pick, 1992; Stucki et al., 1987). The relative motion of the head and facial features may have auditory correlates that distinguish adults from children. Regardless of which hypothesis is ultimately supported, these findings reveal that infants are able to abstract invariant intermodal relations from dynamic displays of faces and voices that differentiate children from adults. The information critical for this matching is contained in upright faces rather than inverted faces.

Further, secondary analyses exploring the role of experience in intermodal matching revealed several interesting trends. In Experiment 1, it appeared that the infant's prior experience with children was important for matching at 7 months but not at 4 months. Infants who had siblings or regular interaction with children showed significant matching of faces and voices, whereas those who had no siblings and rarely interacted with other children showed no matching. Apparently, at 7 months experience with children facilitated abstraction of information uniting the faces and voices on the basis of age. However, at 4 months, although matching was significant by the second block of trials, experience with children was unrelated to this effect. This may have occurred for several reasons. Between 4 and 7 months, infants make a transition from a predominantly reclining position to sitting upright. This enables new exploratory patterns, expands their view of the world, and has consequences for perception (see Gibson, 1988). Thus, although 4-month-olds have had a great deal of exposure to their siblings, they may not attend to the nature of the face-voice relations in the same way that 7-month-olds do. Alternatively, 7-month-olds have had more experience with the faces and voices of their siblings than 4-month-olds, and it may not be until 7 months that the cumulative effects of experience pay off in terms of enhanced face-voice matching. Thus, 7-month-olds catch on more quickly. In either case, our findings that experience enhanced matching of faces and voices in the upright but not the inverted condition suggest that experience with children facilitates abstraction of relational information in the face. Before firm conclusions can be drawn about the role of experience, these trends should be followed up with studies where experience is manipulated a priori and a more sensitive or continuous measure of experience is taken.

Results of Experiment 1 also indicated that infants at 7 months showed a preference for the child faces over adult faces. Four-month-olds showed a marginal preference in the same direction. Infants looked longer at the child faces regardless of the sound condition. Experience with children did not interact with this preference. This preference is consistent with other findings in the literature. Infants prefer to look at children over adults, and show more positive affect toward children and less stranger anxiety (e.g., Bigelow et al., 1990; Brooks & Lewis, 1976; Greenberg et al., 1973; Lewis & Brooks, 1974). However, even with inverted faces, infants showed a significant preference for the child faces over the adult faces.

The robust preference for the child faces in the upright condition may have a number of different bases. For example, adults find the head shape and features of young faces to be "cuter" and more attractive than older faces (Alley, 1981, 1988). Given infants' agreement with adult aesthetics regarding the attractiveness of adult faces (Langlois et al., 1991; Samuels et al., 1994), it is reasonable to hypothesize that they might also perceive attractiveness of young versus older faces in a manner similar to adults. However, the preference for child faces in the inverted condition, where no intermodal matching was observed, reveals that preference and discrimination of children from adults is also possible on the basis of visual information other than facial configuration. Perhaps it is based on something as simple as the smaller, rounder head, smaller features, or greater amount of head and body motion. A parallel pattern of results was found by Kestenbaum and Nelson (1990) in an investigation of infant perception of affective expressions. Infants were able to categorize emotional expressions across different models in an upright but not an inverted condition. However, they discriminated among different expressions posed by a single model in both an upright and an inverted orientation. It was concluded that categorization of affective expressions required attention to relational information (configuration), whereas discrimination could be accomplished on the basis of simple features. Similarly, in our research, infants showed intermodal matching with upright but not inverted faces. Matching requires abstraction of invariants specifying child versus adult faces. However, the preference for the child faces, shown in both the upright and inverted condition, also reveals that discrimination can be based on more simple attributes.

In sum, this research demonstrates infants' ability to relate dynamic faces and voices of strangers on the basis of invariant information that specifies age or maturity. This matching requires an upright face and is thus hypothesized to be based on configurational information, that is, the relative positioning and/or motion of features of the face. As in the case of gender matching, these findings show infants' detection of another complex, typical, intermodal relation where both amodal and arbitrary audiovisual relations unite faces and voices.

ACKNOWLEDGMENTS

This research was supported by National Institutes of Child Health and Human Development grant RO1 HD25669 awarded to the first author. Portions of these data were presented at the 1992 meeting of the International Conference on Infant Studies, Miami, FL, and the 1994 meeting in Paris. The authors are grateful to Linda Alberga, Maricel Cigales, Ivonne Fernandez, Lisa Friedman, Brenda Lundy, and Gloria Peruyera for their assistance with data collection and analyses.

ADDRESSES AND AFFILIATIONS

Corresponding author: Lorraine E. Bahrick, Department of Psychology, Florida International University, Miami, FL 33199; e-mail: bahrick@fiu.edu. Dianelys Netto and Maria Hernandez-Reif are both at the University of Miami.

REFERENCES

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

Alley, T. R. (1981). Head shape and the perception of cuteness. *Developmental Psychology, 17*, 650–654.

Alley, T. R. (1988). Social and applied aspects of face perception: An introduction. In T. R. Alley (Ed.), *Social and applied aspects of perceiving faces*. Hillsdale, NJ: Erlbaum.

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

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

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

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

Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expression. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 373–379.

Bernard, J. A., & Ramey, C. T. (1977). Visual regard of familiar and unfamiliar persons in the first six months of infancy. *Merrill-Palmer Quarterly, 23,* 121–127.

Bertenthal, B. I., Proffitt, D. R., Spetner, N. B., & Thomas, M. A. (1985). The development of infant sensitivity to biomechanical motions. *Child Development, 56,* 531–543.

Bigelow, A., MacLean, J., Wood, C., & Smith, J. (1990). Infants' responses to child and adult strangers: An investigation of height and facial configuration variables. *Infant Behavior and Development, 13,* 21–32.

Biringen, Z. C. (1987). Infant attention to facial expressions and facial motion. *Journal of Genetic Psychology, 148,* 127–133.

Brooks, J., & Lewis, M. (1976). Infants' responses to strangers: Midget, adult, and child. *Child Development, 47,* 323–332.

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

Caron, A. J., Caron, R. F., Caldwell, R. C., & Weiss, S. J. (1973). Infant perception of the structural properties of the face. *Developmental Psychology, 19,* 385–399.

Caron, A. J., Caron, R. F., & MacLean, D. J. (1988). Infant discrimination of naturalistic emotional expressions: The role of face and voice. *Child Development, 59,* 604–616.

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.

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

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

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mother's 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.

Diamond R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115,* 107–117.

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

Enlow, D. H. (1982). *Handbook of facial growth* (2d Ed.). Philadelphia: Saunders.

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

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

Fagan, J. F. (1979). The origins of facial pattern recognition. In M. H. Bornstein & W. Kessen (Eds.), *Psychological development from infancy: Image to intention.* Hillsdale, NJ: Erlbaum.

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

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.

Field, T. M., Woodson, R., Cohen, D., Greenberg, R., Garcia, R., & Collins, K. (1983). Discrimination and imitation of facial expressions by term and preterm neonates. *Infant Behavior and Development, 6,* 485–489.

Fox, R., & McDaniel, C. (1982). The perception of biological motion by human infants. *Science, 218,* 486–487.

Gibson, E. J. (1988). Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology* (Vol. *39,* pp. 1–41). Palo Alto, CA: Annual Reviews, Inc.

Greenberg, D. J., Hillman, D., & Grice, D. (1973). Infant and stranger variables related to stranger anxiety in the first year of life. *Developmental Psychology, 9,* 207–212.

Grieser, D., & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech sound prototypes. *Developmental Psychology, 25,* 577–588.

Kestenbaum, R., & Nelson, C. A. (1990). The recognition and categorization of upright and inverted emotional expressions by 7-month-old infants. *Infant Behavior and Development, 13,* 497–511.

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

Langlois, J. H., Ritter, J. M., Roggman, L. A., & Vaughn, L. S. (1991). Facial diversity and infant preferences for attractive faces. *Developmental Psychology, 27,* 79–84.

Lasky, R. E., Klein, R. E., & Martinez, S. (1974). Age and sex discriminations in 5 and 6 month old infants. *Journal of Psychology, 88,* 317–324.

Lewis M., & Brooks, J. (1974). Self, other, and fear: Infants' reactions to people. In M. Lewis & L. Rosenblum (Eds.), *The origins of fear: The origins of behavior* (Vol. 2). New York: Wiley.

Lewkowicz, D. J. (1986). Developmental changes in infants' bisensory response to synchronous durations. *Infant Behavior and Development, 9,* 335–353.

Lewkowicz, D. J. (1992). Infants' response to temporally based intersensory equivalence: The effect of synchronous sounds on visual preferences for moving stimuli. *Infant Behavior and Development, 15,* 297–323.

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.

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

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.

Nelson, C. A., & Horowitz, F. D. (1983). The perception of facial expressions and stimulus motion by 2- and 5-month-old infants using holographic stimuli. *Child Development, 54,* 868–877.

Nelson, C. A., & Ludemann, P. M. (1989). Past, current, and future trends in infant face perception research. *Canadian Journal of Psychology, 43,* 183–198.

Oster, H. (1981). Recognition of emotional expressions in infancy? In M. E. Lamb & L. R. Sherrod (Eds.), *Infant social cognition: Empirical and theoretical considerations.* Hillsdale, NJ: Erlbaum.

Pickens, J. N., & Bahrick, L. E. (1995). Infants' discrimination of bimodal events on the basis of rhythm and tempo. *British Journal of Developmental Psychology, 13,* 223–236.

Pickens, J. N., & Bahrick, L. E. (1997). Do infants perceive invariant tempo and rhythm in auditory-visual events? *Infant Behavior and Development, 20,* 349–357.

Poulin-Dubois, D., Serbin, L. A., Kenyon, B., & Derbyshire, A. (1994). Infants' intermodal knowledge about gender. *Developmental Psychology, 30,* 436–441.

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.

Samuels, C. A., Butterworth, G., Roberts, T., Graupner, L., & Hole, G. (1994). Facial aesthetics: Babies prefer attractiveness to symmetry. *Perception, 23,* 823–831.

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

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

Stucki, M., Kaufmann-Hayoz, R., & Kaufmann, F. (1987). Infants' recognition of a face revealed through motion: Contribution of internal facial movement and head movement. *Journal of Experimental Child Psychology, 44,* 80–91.

Trehub, S. E., & Thorpe L. A. (1989). Infants' perception of rhythm: Categorization of auditory sequences by temporal structure. *Canadian Journal of Psychology, 43,* 217–229.

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

Walker-Andrews, A. S. (1988). Infants' perception of the affordances of expressive behaviors. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (Vol. 5, pp. 173–221). Norwood, NJ: Ablex.

Walker-Andrews A. S. (1994). In D. 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). Infant's bimodal perception of gender. *Ecological Psychology, 3,* 55–75.

Walton, G. E., Bower, N. J., & Bower, T. G. (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.