ECOLOGICAL PSYCHOLOGY, 3(2), 55-75

Infants' Bimodal Perception of Gender

Arlene S. Walker-Andrews
Rutgers University

Lorraine E. Bahrick
Florida International University

Stacy S. Raglioni
Rutgers University

Isabel Diaz
Florida International University

Two experiments were independently conducted in separate labs to determine whether infants are sensitive to intermodal information specifying gender across dynamic displays of faces and voices. In one study, 4- and 6-month-old infants were presented simultaneously with a single videotape of a male face and a female face accompanied by a single voice for two 2 min trials. In the second study 3 ½ and 6 ½ month olds were also presented videotapes of male and female faces accompanied by a single voice but for a series of short trials. Temporal synchrony between face and voice was controlled in both studies by presenting both male and female faces speaking in synchrony with a single soundtrack. In both experiments the 6 month olds showed evidence of matching faces and voices on the basis of gender. They significantly increased their looking to a face when the gender-appropriate voice was played. Four month olds gave evidence for matching the faces and voices based on gender information only on the second trial of Experiment 1, whereas the 3 ½ month olds failed to show any preferential looking.

Studies of infants' intermodal perception have continued to appear in the literature, perhaps because they hold promise for distinguishing between several perspectives on the development of perception. One unresolved question concerns whether infants' sense modalities are integrated or differentiated at birth. According to an integration-association view, sensations from different recep-

Requests for reprints should be sent to Arlene S. Walker-Andrews, Psychology Department, Tillett Hall, Rutgers University, New Brunswick, NJ 08903.

tors are separate at birth and the developmental task is to coordinate them to achieve a single percept (e.g., Birch & Lefford, 1963; Bryant, 1974; Piaget, 1954). Another view, attributed primarily to Bower (1974), asserts that the newborn infant is born with undifferentiated perceptual systems and, with development, differentiation occurs. Initially, there is a *primitive unity*, which means that the infant is unaware even of whether an object is seen, heard, or felt (Bower, 1974). Our view, drawn from the theories of J.J. Gibson and E.J. Gibson (E.J. Gibson, 1983), holds that detection of invariant relations underlies the development of perception. It emphasizes that objects and events in the world are specified by redundant information abstracted through several sense modalities: "The five perceptual systems correspond to five modes of overt attention. They have overlapping functions, and they are all more or less subordinated to an overall orienting system" (J.J. Gibson, 1979/1986, p. 245). To focus entirely on the question of separability or unity is to neglect that the "achievements of the perceptual systems are specific to the qualities of things in the world, especially their affordances" (J.J. Gibson, 1979/1986, p. 246). From this view, our task is to determine to which intermodal correspondences infants are sensitive, delineate the limits and mechanisms underlying such sensitivity, and discover how these mechanisms develop. Insofar as an infant detects amodal invariant information over time, his or her perceptions will be of objects and events and their potential affordances rather than separate sensory impressions. We present data here from infants of 3 ½, 4, and 6 months old that speak to their ability to detect a match between same gender faces and voices.

Infants are remarkably sensitive to amodal invariant relations: They detect temporal synchrony, tempo of action, rhythm, changing distance, and affective information uniting visual and acoustic presentations (e.g., Bahrick, 1983, 1988; Dodd, 1979; Mendelson & Ferland, 1982; Spelke, 1979; Walker, 1982; Walker-Andrews & Lennon, 1985). Some limitations on infants' abilities have been discovered as well. For example, Spelke, Born, and Chu (1983) found that infants matched impact sounds with either the sight of an object hitting against a surface or a simple change in trajectory of motion with no visual impact. The developmental pattern seems to be toward increasing specificity, because adults in the study required that a sound co-occur with a visual impact. Thus, perceptual learning may lead to refinement of the detection of invariants.

In this set of experiments, we examined infants' perception of information for gender across the visual and auditory modalities. Gender is a naturally occurring class to which infants typically have been exposed. Information specifying gender is available across both the visual and acoustic modalities, but the relationships between face and voice are difficult to delineate. We can identify faces as male or female most of the time, but it is hard to judge what information is critical. Likely candidates include size of features, perceived height of cheekbones (see Fagan & Singer, 1979), as well as culturally specific variations such as facial hair or hair length. For voices, men's tend to be lower pitched, although

there is overlap. In addition, vowels produced by a man are different in formant frequencies from those produced by a woman or a child (Peterson & Barney, 1952). These variations are perceived primarily as a difference in timbre. Such differences result primarily from the length of the supralaryngeal vocal tract and the size of the vocal cords. (For a discussion of the anatomy and physiology of speech, see Lieberman & Blumstein, 1988; Tartter, 1986.) These may also have visible correlates: The human male adult tends to be larger—he has a longer and bigger throat, protuberant Adam's apple, larger face, and larger chest. Thus infants could learn to match same gender faces and voices by (a) detecting invariant audio–visual relations, (b) learning to associate modality-specific information and generalizing to gender categories, or (c) both. In any event, it first must be established whether infants match faces and voices according to gender, and if so, at what age this ability appears.

In general, research about infants' perception of gender information has shown that by about 7 months, infants perceive similarities among same gender faces and same gender voices. Cohen and Strauss (1979) found that infants at 30 weeks, but not younger, showed a transfer of habituation to both a female face they had seen during a habituation sequence and to a novel female face (see also Cornell, 1974; Fagan, 1976, 1979; Fagan & Singer, 1979; Leinbach & Fagot, 1986). With voices, Miller, Younger, and Morse (1982) found that 7 month olds could discriminate between male and female exemplars. Miller (1983) further demonstrated categorization of voices by gender. In a habituation procedure, infants as young as 6 months showed renewed interest to a voice from a new gender category but not to one of the same gender category.

A few studies have also assessed the infants' ability to relate specific face and voice information for male and female actors. Spelke and Owsley (1979) tested infants in an intermodal preference procedure using the faces and voices of the infants' mothers and fathers. Infants of 3 ½ to 7 ½ months looked preferentially to the parent whose voice was heard. It cannot be determined from this study, however, whether infants were using information related to gender or whether intermodal matching was based on more idiosyncratic information learned through extensive experience with their particular parents.

To address this issue, Miller and Horowitz (1980) presented 8 month olds with paired slides of unfamiliar male and female faces along with male or female taped voices. Although subjects tended to look longer at a face when the gender-appropriate voice was played, this matching effect held only for the male voices and faces. Earlier, Lasky, Klein, and Martinez (1974) had tested 5- and 6-month-old infants with paired photographs of a man and woman, a man and a boy, or a woman and a boy, accompanied by each voice of the pair in sequence. Infants looked longer at the male photograph and showed a visual preference for the voice-appropriate photographs only when the woman was paired with the boy's face. Together, these findings provide little evidence for intermodal knowledge of faces and voices according to gender categories. Furthermore, in both studies,

use of a static face with a voice precluded intermodal matching on the basis of dynamic relations.

Francis and McCroy (1983) investigated infants' sensitivity to dynamic, bimodal presentations of male and female adults. Infants of 3, 6, and 9 months were shown a male face and a female face accompanied by a female voice, a male voice, or music. However, two separate video monitors were used for the visual presentations, making it unlikely that the two faces and accompanying voice could have been precisely synchronized throughout. Thus, the role played by voice–face synchrony could not be determined. Results revealed that 6-month-old infants who heard the male voices looked longer at the male faces compared with infants who heard music or female voices. Infants who heard the female voices looked longer at the female faces compared to infants who heard male voices but not significantly more than infants who heard music. In contrast, the 3 month olds looked longer to the female face regardless of the auditory condition, whereas the 9 month old exhibited no significant looking preferences at all. These mixed results allowed no firm conclusions.

In conclusion, little is known about when and under what conditions infants are able to relate dynamic faces and voices based on gender. Furthermore, research using dynamic displays requires ruling out bases of face–voice matching other than gender, such as audio–visual synchrony, tempo of action, and amount and type of affect. Because infants as young as 4 months can match visual and acoustic information using both of these temporal invariants (e.g., Bahrick, 1983, 1988; Dodd, 1979; Spelke, 1979, 1981; Walker, 1982), as well as affective expression (Walker, 1982; Walker-Andrews, 1986), special attention must be given to equating the male and female voice–face displays for these variables.

Our research asked whether young infants could detect intermodal relations in dynamic male and female faces and voices, while eliminating the potential confounds just mentioned. Furthermore, given the importance of independent replication with different stimulus materials, experimenters, and procedures in infant research, two studies were carried out independently by different research labs. The first and second authors agreed on the research question, ages to be tested, and the importance of the aforementioned controls and otherwise worked independently. The age groups were selected to assess our hypothesis that matching based on gender information would develop between 3 and 6 months. The resulting studies reflect the ecological diversity normally observed in separate manuscripts, yet the discussion benefits from converging findings. Each study assessed whether 3- to 4- and 6-month-old infants could detect intermodal relations uniting the faces and voices of men and women in dynamic, carefully controlled presentations. Two similar methods were used. In both studies infants were shown videotapes of a male and female speaking side by side along with a single soundtrack corresponding to one of them but synchronized with the motions of both. Tempo of action and temporal

synchrony between the voice and lip motions were controlled by presenting the auditory passage in synchrony with the motions of both the male and female actors simultaneously. In addition, special attention was devoted to equating the male and female actors for both facial and vocal affect. Given these controls, the infants' ability to match male faces with male voices and female faces with female voices suggested that the infant is capable of detecting information specifying gender across the visual and acoustic modalities.

Differences between the two studies included length of stimulus presentation, number and sequence of voice-accompanied trials, auditory passages, and the male and female models. By using different pairs of male and female actors, different variations of the intermodal preference procedure, and two labs, we could test the generality of any intermodal matching across stimuli, procedures, and context. In addition, by including the results in a single article, we had the unusual opportunity of asking the same questions with two sets of data collected from different samples.

EXPERIMENT 1

This study was conducted under the supervision of Arlene Walker-Andrews.

Method

Infant observers. Sixteen 4 month olds (8 boys and 8 girls with a *M* age of 129.4 days, *SD* = 9.1 days) and sixteen 6 month olds (8 boys and 8 girls with a *M* age of 179.1 days, *SD* = 8.23 days) participated. Data from 11 additional infants (8 4 month olds and 3 6 month olds) were collected but eliminated because of experimenter error or excessive fussing.

Stimulus materials. Color videotapes were used in this experiment. Each videotape depicted both a male and female adult standing side by side, speaking continuously with neutral facial expressions. Each person recited *The Twelve Days of Christmas* for 2 min. Both individuals were White, had dark brown hair, and blue eyes. The man's hair covered the tips of his ears; the woman's was chin length. A Panasonic video camera (Type WV-3180) was used to record the actors, including a view of the face and shoulders only. Synchrony between the face and lip movements was accomplished by having the actors stand together and recite the poem simultaneously. One videotape was made of the man standing to the right of the woman and one videotape of the man standing to the left of the woman. Copies were made of each videotape and then each actor dubbed his or her voice onto two of the four copies (man left, male voice dubbed; man left, female voice; man right, male voice; man right, female voice). The actors were instructed to maintain a neutral face and voice throughout the

videotaping and audio-dubbing. Four adult observers who viewed each actor–voice combination (while the other actor was masked from view) reported that although the female voice seemed odd coming from the male actor (and vice versa) asynchrony was not evident.

Apparatus and procedure. Infants were seated in a standard infant seat about 80 cm from the display. A flashlight centered above the display attracted the infant's visual attention before each presentation. An aperture below or to the left (when there were two observers) of the display permitted a trained observer to monitor an infant's fixations. Although the observer could hear each soundtrack, she was unaware of the lateral positions of the male and female actors; she depressed buttons connected to an event recorder to indicate whether an infant was looking to the right or left. Fixations judged off screen or at center were not recorded. Interobserver reliability for 25% of the infants averaged .96 (see Bahrick, Walker, & Neisser, 1981).

The videotapes were presented on a 19″ color monitor (Panasonic Model No. BT–S1900N; video cassette recorder Panasonic NV–8350). A large Fresnel lens (73.7 × 54.6 cm) was placed in front of the monitor to magnify the videotaped image of the two faces.[1] The soundtrack for each videotape was played through a speaker above the images. The soundtrack averaged 50 to 55 dB at the infants' position, although there were variations in amplitude for a voice during a trial. Therefore, each infant viewed a single videotape during a trial so that synchrony between the actors could be maintained, but the videotaped image was magnified so that the faces were nearer life-size and so that an observer could monitor fixations more accurately.

Infants viewed two videotapes, each of a male and female actor side by side, for two trials. On the first trial, the film was accompanied by one soundtrack (e.g., male voice), and (after a 2 to 3 s interval) on the second trial it was accompanied by the other soundtrack (e.g., female voice). Each of the trials lasted 2 min. Order of soundtrack and lateral position of the first sound-specified individual were counterbalanced across subjects. For half the infants, the lateral position of the male and female actors was switched from Trial 1 to Trial 2, whereas for the other half, it remained constant.

Results and Discussion

Duration of fixation to each actor (fixation to the left, fixation to the right) was recorded for each infant. These fixation times were expressed as the proportion of total looking time (PTLT; number of seconds fixation to one actor divided by number of seconds fixation to both) and included in a number of analyses. The

[1]Because of this setup the infants were actually looking at a virtual image which subtended a visual angle of 18°.

primary measure of interest was the difference in PTLT to the female actor when accompanied by the gender-matching voice as contrasted with the mismatching (male) voice. Because these proportions are taken from two different trials, they are independent and can be statistically compared. Therefore, difference scores were also calculated for each infant by subtracting the PTLT to the female face when the voice was mismatched from that to the female face when the voice was gender-matched. (Difference scores for the male face when it was sound-specified vs. mismatched are the reciprocal of that derived for females and thus are not discussed further.) These results are presented in Table 1 and show fixation time (in seconds), PTLT (for sound-matched vs. mismatched films) and difference scores for infants at each age. Analyses conducted on these and other measures are given next.

Overall the infants (4 and 6 month olds combined) looked to the films about 70% of the time available. On Trial 1 TLT averaged 88.33 s (SD = 20.79) and on Trial 2, 80.71 s (SD = 20.82). Fixation, although high throughout, declined across the two 2 min trials, $t(31) = 2.07$, $p = .047$, two-tailed.[2]

A 2 × 2 × 2 analysis of variance (ANOVA) was conducted comparing the difference scores for the 4- and 6-month-old infants to determine whether there were any differences in looking attributable to the age of infant, sex of infant, or to whether the sound-specified film occurred on the same or different sides on each trial. There were no significant main effects for age, $F(1, 24) = .19$, $p = .666$; sex, $F(1, 24) = .74$, $p = .397$; side change or not, $F(1, 24) = 1.29$, $p = .266$. The interaction of age and sex approached significance, $F(1, 24) = 3.46$, $p = .075$, as the difference scores for 4-month-old boys were at zero (−.003), compared to a mean of .254 for the remaining Age × Sex groups. No other interactions were significant for Age × Side Change, $F(1, 24) = .00$, $p = .983$; Sex × Side Change, $F(1, 24) = .04$, $p = .843$; Age × Side Change × Sex, $F(1, 24) = .00$, $p = .946$.

To address our primary question, whether infants demonstrate audiovisual matching based on information for gender and at what age, PTLT's and difference scores were examined separately for each age group. The 6 month olds looked differentially to the videotaped actors. They markedly increased their looking time to a particular person when that person's voice was heard, compared with when the other gender voice was played. The average PTLT directed at the female face when it was voice mismatched was .410 and .575 when it was accompanied by the matching voice, $t(15) = 2.32$, $p < .05$. On average, 6 month olds spent a significant PTLT (M = .582) fixating the sound-matched films, $t(15) = 2.31$, $p < .05$, according to a single-sample t test against the chance value of .50. At the individual subject level, 12 of the 16 infants showed a greater PTLT to the sound-matched film than the mismatched film. This is significant ($p < .05$) according to a binomial test. Thus, the 6

[2]All statistical tests reported were two-tailed.

TABLE 1
Seconds of Looking, Proportions of Total Looking Time (PTLT), and Difference Scores to Visual Displays of Male and Female Faces

			Fixation By Sound Manipulation			*Fixation By Gender & Sound Manipulation*			
			Sound Matched	*Sound Mismatched*	*Difference*	*Fm*		*Mf*	
Months	*Trial*		*F + M*	*f + m*	*F − f*	*F*	*m*	*M*	*f*
4	T1	$\bar{X}$sec	97.6	84.53		43.4	45.9	54.2	38.63
		$\bar{X}$PTLT	.526[a]	.474	+0.53	.488	.512	.567	.434
		SD	(.29)	(.29)	(.58)	(.28)	(.28)	(.31)	(.31)
		N	16	16					
		n				8	8	8	8
	T2	$\bar{X}$sec	104.02	59.90		40.49	39.61	63.53	20.29
		$\bar{X}$PTLT	.636	.363	+.273	.525	.475	.748	.252
		SD	(.25)	(.25)	(.50)	(.30)	(.30)	(.14)	(.14)
		N	16	16					
		n				8	8	8	8
	T1 + T2	$\bar{X}$sec	100.82	72.22		41.94	42.76	58.88	29.46
		$\bar{X}$PTLT	.582	.419	+.163	.506	.494	.657	.343
		SD	(.21)	(.21)	(.42)	(.28)	(.28)	(.25)	(.25)
		N	16	16					
		n				8	8	8	8

6	T1	$\bar{X}$sec	93.32	77.85		40.38	36.34	52.94	41.51
		$\bar{X}$PTLT	.523	.476	+.046	.489	.510	.557	.443
		SD	(.22)	(.22)	(.44)	(.23)	(.23)	(.22)	(.22)
		N	16	16					
		n				8	8	8	8
	T2	$\bar{X}$sec	101.67	57.26		52.68	30.26	48.99	27.0
		$\bar{X}$PTLT	.642	.359	+.283	.661	.339	.622	.378
		SD	(.20)	(.20)	(.41)	(.20)	(.20)	(.22)	(.22)
		N	16	16					
		n				8	8	8	8
	T1 + T2	$\bar{X}$sec	97.49	67.56		46.53	33.30	50.96	34.26
		$\bar{X}$PTLT	.582	.481	+.165	.575	.425	.589	.410
		SD	(.14)	(.14)	(.29)	(.23)	(.23)	(.21)	(.21)
		N	16	16					
		n				8	8	8	8

Note. A visual display which is matched by the soundtrack is designated with a capital "F" or "M;" the mismatched display is designated by a lower-case "f" or "m." Side of presentation was counterbalanced across subjects.

[a]Looking preferences were computed by dividing the number of seconds each subject looked at the sound film by the total time spent looking at both films separately and averaging across subjects. The average proportions (F + M, f + m) differ from the proportions derived from the raw number of seconds (Xsec) and from averaging the proportions to F and M because the mean of ratios is not, in general, equal to the ratio of means.

month olds looked longer at matching dynamic displays of male faces and voices and female faces and voices both across trials and at the individual level.

Within trials, the 6 month olds showed a significant PTLT to the sound-matched film on Trial 2, $M = .642$, $t(15) = 2.76$, $p < .05$; this proportion was not significant for Trial 1 alone, $M = .523$, $t(15) = .409$, $p > .10$. Further analyses assessed preferences for the male face versus the female face and for lateral position. Independent of the sound manipulation, infants did not show preferential looking at the male or female faces—mean PTLT to the male face was .518, $t(15) = .70$, $p > .10$. They did, however, show a nonsignificant tendency to look to the right side of the screen, $M = .579$, $t(15) = 1.93$, $p < .10$.

In contrast, the 4-month-old infants did not demonstrate intermodal matching across trials. Although these infants tended to increase their looking time to a particular actor when that person's voice was played, compared with when the other gender voice was played, that increase was not significant. The average PTLT directed to the female face when it was voice mismatched was .343 and when it was voice matched .506, $t(15) = 1.55$, $p < .10$. Overall, the proportion of looking time to the sound-matched face averaged .581, $t(15) = 1.55$, $p < .10$. Eleven of the 16 infants looked longer than 50% of the time to a face when the gender-appropriate voice was played ($p > .05$ according to a binomial test).

Within trials, the PTLT to the sound-matched face was greater on Trial 2 ($M = .636$) than Trial 1 ($M = .526$). The PTLT for Trial 1 was not different from chance, $t(15) = .365$, $p > .10$, whereas that of Trial 2 alone exceeded .50, $t(15) = 2.14$, $p < .05$. Four month olds thus demonstrated matching based on gender information for Trial 2 alone. Further analyses assessed whether the 4 month olds showed evidence of a preference for the male or female face or a lateral position, independent of the sound manipulation. Results indicated no significant preference for the male or female face, $M = .425$, $t(15) = 1.82$, $p > .10$, or for one side of the screen over the other—right side, $M = .536$, $t(15) = .59$, $p > .10$, according to single-sample t tests against the chance value of .50.

Overall, 6 month olds showed matching based on information for gender across both trials according to the PTLT and difference score measures, as well as at the individual subject level. Four month olds gave no evidence of matching across trials but did show a matching effect for Trial 2 alone. Perhaps it took these younger infants more time to abstract the intermodal relations, enabling them to match faces and voices by the second half of the procedure.

EXPERIMENT 2

This study was conducted under the supervision of Lorraine E. Bahrick. Because this study and Experiment 1 were independent replications, many details of the stimulus displays, apparatus, and procedures differed, as well as the reporting of

results. These variations across studies were preserved to enhance external validity of the findings should they be parallel.

Separate stimulus films were developed for this study with new male and female actors and auditory passages. In addition, the methods for achieving synchrony between the motions of the male and female video displays differed, as did the number, length, and presentation format of the preference trials. Key aspects of the procedure, however, were replicated from one study to the next. That is, videotaped images depicting a male and female actor were presented side by side, along with a single soundtrack that was synchronized with the lip motions of both actors but was gender appropriate to only one of them.

Method

Infant observers. Twenty-four 3 ½ month olds (14 girls and 10 boys, with a M age of 106.6 days, *SD* = 5.8 days) and twenty-four 6 ½ month olds (14 girls and 10 boys, with a M age of 195.8, *SD* = 4.3 days) participated. Data from 14 additional infants (ten 3 ½ month olds and four 6 month olds) were collected and eliminated from the study because of experimenter error ($n = 7$) or excessive fussiness ($n = 7$).

Stimulus materials. Two sets of color videotapes were made, each depicting a man and a woman reciting a nursery rhyme, *This Old Man*, for 2 min at about 55 dB. One stimulus set depicted a man and woman, both with light hair, blue eyes, and fair coloring, and the other a man and woman, both with dark brown hair, brown eyes, and darker coloring. Both men had fairly short hair, and both women had shoulder length hair. A Panasonic video camera (Type WV-3170) was used to record the actors, including a view of the face and shoulders only. Synchrony between the face and lip motions of the male and female actors was accomplished by having each actor synchronize his or her speech and lip motions with that of a videotaped model. Furthermore, to ensure similar affect across individuals, the actors also mimicked the facial expressions of the model, specifically rehearsing the timing of all smiles and eyebrow movements. Then, to ensure that the degree of face–voice synchrony was similar across conditions, each actor dubbed his or her voice onto the final video image of his or her speaking face. Finally, as a double check to determine whether matching of faces and voices could be performed based on affect or amount of animation, adult judges rated the moving faces ($n = 6$) and the voices ($n = 5$) according to how happy and animated they seemed. Judges showed no ability to match faces and voices on the basis of these aspects. In fact, most judges (four out of five) agreed that although the moving face of the male brunette seemed happier and more animated than that of the female brunette, none thought that his voice sounded happier or more animated; rather, most thought that the voice of the woman was more animated and happy sounding. A mismatching pattern was also found

for the blond pair. Most judges agreed that the female face seemed happier and more animated than the male face, although none thought that her voice sounded happier. Thus, we were confident that any evidence of matching on the part of infants would not be a result of matching on the basis of face–voice animation or affect.

Apparatus and procedure. Infants were seated in a standard infant seat facing two color video monitors (Panasonic BT–S1900N) about 50 cm away. A column of colored lights and a small mechanical toy dog were positioned between the monitors to attract the infant's gaze between trials. Black posterboard enclosed the video monitors. Two apertures, one between them and the other to the left of center, were cut into the posterboard from which observers could monitor the infants' visual fixations. A trained observer, blind to the lateral positions of the display, monitored visual fixations by depressing one of a set of two buttons to indicate whether the infant was fixating the right- or left-hand display. A second observer monitored visual fixations for 31 of the subjects. Interobserver reliability, expressed as a Pearson product–moment correlation between the looking proportions derived from the primary versus secondary observer was .97.

All stimulus displays were presented via one of two Panasonic video decks (NV–8500 and AG–6300) that were connected to an edit controller (Panasonic NV–A500). The edit controller allowed us to precisely synchronize the output from two video films (to the nearest .02 s), so that the lip movements of the male and female speaker could be kept aligned throughout the stimulus presentation. All soundtracks emanated from a speaker located midway between and just below the two video screens.

Infants viewed sixteen 20 s trials comprised of two blocks of 8 trials. Each block was identical for an infant allowing for comparison of an infant's performance from one block to the other because the procedure was lengthy. Each trial depicted a man and a woman, side by side, speaking in synchrony with one another, along with the synchronized soundtrack appropriate to one of them. The order of soundtrack presentation (male voice, female voice) was semi-random, with the restriction that each infant received four presentations of each voice in each trial block and that a particular soundtrack was played no more than twice in succession. A given random soundtrack order was selected for each infant for the first block of 8 trials and this order was then repeated for the second block of 8 trials. The intertrial interval was approximately 4 s long and the infant's visual attention was attracted to center during this interval by the flashing lights and toy dog. Half of the infants of each age group received the stimulus set depicting the dark-haired actors, whereas the other half received the set depicting the light-haired actors. The lateral positions of the male and female video displays were counterbalanced across infants with each stimulus set condition. Half received the man on the right and the woman on the left

throughout the 16-trial session, whereas the other half received the opposite arrangement. Thus, although the position of the male and female visual display remained constant across trials for a given subject, the male and female auditory displays varied randomly across trials.

Results and Discussion

Overall, infants looked to the films about 205 s out of the total 320 s (64% of the time). TLT for the first block of 8 trials was 110.42 s and for the second block of 8 trials was 95.47 s. Visual fixations were expressed as the PTLT infants fixated the sound-matched display on each trial, and these proportions were averaged across each trial block (Trials 1 to 8 and Trials 9 to 16) to obtain a mean PTLT to the sound-matched display. A grand mean PTLT was also derived by averaging across the two blocks for each subject. Note that the two trial blocks represent essentially two independent replications of the procedure. Furthermore, as in Experiment 1, difference scores were calculated for each trial block reflecting the PTLT to the female actress when the gender-matching voice was played minus the PTLT to the female actress when the mismatching male voice was played. As described for Experiment 1, positive difference scores indicate greater looking to the voice-matched face whereas negative scores indicate greater looking to the voice-mismatched face. Results of the study are depicted in Table 2 and parallel the format given for Experiment 1. The table displays the seconds fixation time and PTLT for sound-matched versus mismatched films, as well as difference scores for each block of trials at each age.

Observers reported that subjects seemed fatigued and more fussy toward the end of the session (possibly because of the numerous trials, see Lasky et al., 1974). Subsequent analyses confirmed our observations. Both the 3 ½ and 6 ½ month olds showed a decrease in total looking to the video displays during Trials 9 to 16 as compared with Trials 1 to 8. Six month olds looked an average of 99.5 s (62% of the time available) during Trials 1 to 8 and 81.9 s (51%) during Trials 9 to 16, $t(23) = 4.55$, $p < .0001$; 3 month olds looked an average of 119.3 s (75% of the time available) during Trials 1 to 8 and 109.1 s (68%) during Trials 9 to 16, $t(23) = 1.95$, $p < .10$.

To assess any differences across age and trials in looking to the sound-matched film, a two-way ANOVA with age as a between-subjects factor and trial block (1 to 8; 9 to 16) as a within-subjects factor was conducted on the looking proportions. There were no significant main effects of age, $F(1, 46) = .58$, $p > .1$, or trial block, $F(1, 46) = 2.04$, $p > .1$, or an interaction, $F(1, 46) = .46$, $p > .1$. Thus, infants did not differ across age or trial block in their tendency to watch the gender-specified display. Given that the effect of trial block was not significant and because subjects were significantly less attentive during the second half of the study, for subsequent analyses, more importance is

TABLE 2
Seconds of Looking, Proportions of Total Looking Time (PTLT), and Difference Scores to Visual Displays of Male and Female Faces

			Fixation By Sound Manipulation			*Fixation By Gender & Sound Manipulation*			
			Sound Matched	*Sound Mismatched*	*Difference*	*Fm*		*Mf*	
Months	*Trial*		F + M	*f* + *m*	F − *f*	F	*m*	M	*f*
3½	1 to 8	$\bar{X}$sec	61.87	57.44		34.92	25.05	26.95	32.39
		$\bar{X}$PTLT	.512[a]	.488	+.025	.592	.408	.433	.567
		SD	(.11)	(.11)	(.22)	(.31)	(.31)	(.36)	(.36)
	9 to 16	$\bar{X}$sec	54.49	54.59		29.78	24.88	24.71	29.71
		$\bar{X}$PTLT	.495	.505	−.01	.545	.455	.445	.555
		SD	(.08)	(.08)	(.16)	(.30)	(.30)	(.27)	(.27)
	1 to 16	$\bar{X}$sec	116.36	112.03		64.70	49.93	51.66	62.10
		$\bar{X}$PTLT	.502	.489	+.005	.566	.434	.438	.561
		SD	(.07)	(.07)	(.14)	(.29)	(.29)	(.28)	(.28)
6	1 to 8	$\bar{X}$sec	54.36	45.17		26.53	23.75	27.83	21 42
		$\bar{X}$PTLT	.540	.460	+.08	.529	.471	.551	.449
		SD	(.09)	(.09)	(.18)	(.17)	(.17)	(.20)	(.20)
	9 to 16	$\bar{X}$sec	43.00	38.86		21.26	20.04	21.74	18.82
		$\bar{X}$PTLT	.506	.494	+.012	.475	.525	.537	.463
		SD	(.10)	(.10)	(.20)	(.18)	(.18)	(.17)	(.17)
	1 to 16	$\bar{X}$sec	97.36	84.03		47.79	43.79	49.57	40.24
		$\bar{X}$PTLT	.522	.478	+.043	.500	.500	.543	.457
		SD	(.08)	(.08)	(.15)	(.14)	(.14)	(.17)	(.17)

Note. A visual display which is matched by the soundtrack is designated with a capital "F" or "M;" the mismatched display is designated by a lower-case "f" or "m." (N = 24 each cell).

[a]Looking preferences were computed by dividing the number of seconds each subject looked at the sound film by the total time spent looking at both films separately and averaging across trials and subjects. The average proportions (F + M, f + m) differ from the proportions derived from the raw number of seconds (Xsec) and from averaging the proportions to F and M because the mean of ratios is not, in general, equal to the ratio of means.

given to results derived from the first block of 8 trials as opposed to the second block of 8 trials.

To address the primary research question, that is, whether infants showed evidence of matching based on gender information at either age, *t* tests were conducted separately for each age and each block of trials on the PTLT directed toward each face when it was sound-matched versus mismatched (difference scores). Six month olds showed significant evidence of matching on the first block of 8 trials, $t(23) = 2.16$, $p < .05$. Their PTLT to the male face averaged .551 when the male voice was played and only .471 when the female voice was played. Conversely, PTLT to the female face averaged .529 when her voice was played and .449 when the male voice was played. Six month olds showed a significant increase in looking to the male and female faces when the gender-matched voices were played. This effect was not significant during the second block of 8 trials, $t(23) = .30$, $p > .10$. On average 6 month olds spent a significant PTLT fixating the sound-matched films during Trials 1 to 8, $M = .540$, $t(23) = 2.17$, $p < .05$, but showed no significant evidence of matching during Trials 9 to 16, $M = .506$, $t(23) = .06$, $p > .10$.

At the individual level, for the first block of 8 trials, 16 of the twenty-four 6 month olds showed greater looking to the sound-matched face than to the sound-mismatched face. According to a binomial test this was significantly greater than chance ($p < .05$). A significant number of infants also showed greater looking to the sound-matched face across all 16 trials averaged (17 of 24, $p < .05$) whereas results were nonsignificant for the second block of 8 trials taken alone (13 of the 24 infants; $p > .10$). These results parallel those of the group means showing significant matching for the first block of trials and attenuated effects for the second block. Thus, the 6 ½ month olds gave evidence of matching based on gender information during the first block of trials for the PTLT, difference score, and individual subject measures.

The 3 ½ month olds, on the other hand, did not significantly increase their looking time to the sound-matched film on either block of trials, $t(23) = .55$, $p > .10$, Trials 1 to 8; $t(23) = -.29$, $p > .10$, Trials 9 to 16, giving no evidence of matching based on gender information. Their PTLT to the sound-matched film on Trials 1 to 8 averaged .512, $t(23) = .54$, $p > .10$, and on Trials 9 to 16 averaged .495, $t(23) = -.11$, $p > .10$, showing no matching effect according to single sample *t* tests against the chance value of .50. At the individual level, only 10 of the 24 infants showed greater looking to a face when it was sound-matched versus mismatched (13 showed less fixation and 1 showed equal amounts to both), showing no significance according to a binomial test. Thus, the 3 ½ month olds showed no matching on the basis of gender information on either trial block or at the individual subject level.

Further analyses were completed to determine whether infants at either age showed any preferences for the male or female faces or the right- or left-hand displays independent of the sound manipulation. There were no significant

preferences for the male or female face at either 3 ½ or 6 months, $t(23) = 1.06$, $p > .05$; $t(23) = .82$, $p > .05$, respectively, and no significant side preference at 3 ½ or 6 ½ months, $t(23) = 1.26$, $p > .05$; $t(23) = .34$, $p > .05$, respectively. Additional analyses were conducted to determine whether looking proportions to the sound-matched film differed as a function of the stimulus set each infant received. A two-way ANOVA with age and stimulus set as main factors revealed no significant main effects or interaction ($p > .10$, all effects).

In summary, consistent with the results of Experiment 1, 6 ½-month-old infants spent a significantly greater proportion of their visual fixation time viewing male and female faces when gender-appropriate voices were played. They did this even though temporal synchrony relations united both the male and female visual displays with the soundtrack. In contrast, the 3 ½ month olds showed no evidence of intermodal matching based on gender information.

GENERAL DISCUSSION

Results of Experiments 1 and 2 provide converging evidence that 6-month-old infants are capable of detecting intermodal relations specifying speaker gender. The two experiments were conducted in different research labs with different experimenters, different stimulus displays, and two variants of the intermodal visual preference method. In both studies, 6 month olds increased their looking time to a face when a gender-appropriate voice was played, even though both visual displays shared a temporal synchrony relation with the soundtrack and were equally displaced from the source of sound. Furthermore, a significant number of subjects showed this matching effect in both studies. These separate findings increase the generalizability of our conclusions: The effects were not specific to any particular stimulus displays or procedural variations. In contrast, the youngest infants did not provide strong evidence of detecting correspondences between the same gender faces and voices: at 3 ½ months there was no evidence of matching. At 4 months there was a significant tendency to look at a face when the same gender voice was played but only on the second trial.

There were several procedural differences between Experiments 1 and 2 that could have contributed to the different patterns of matching observed in the younger age groups. The procedure used in Experiment 1 may have been less demanding of infants' attentional capabilities. For example, in Experiment 1 infants viewed two individuals accompanied by a single voice for the duration of the first 2 min trial. On the second trial, the voice changed, and the side of presentation for a particular individual changed for half of the infants. In essence, the task for the infant was to select immediately the appropriate individual and continue to view that individual for the duration of the 2 min trial, ignoring his or her lateral position. This procedure sometimes leads to a preference in infants for one side of the viewing screen. In support of this notion,

the 6 month olds tended to look longer at the right-hand screen, although the tendency did not reach significance. In contrast, the procedure used in Experiment 2 was devised initially to reduce infants' tendency to continue looking at a given side or stimulus film (Bahrick, 1983, 1988; Spelke, 1979). Therefore, in Experiment 2, infants viewed two individuals always in the same position but with the voice changing randomly from one trial to the next across the 2 sets of 8 trials. The task for the infant was to find the appropriate individual on hearing a change in voice. The shorter trials and requirement that an infant re-orient to center before onset of each trial would attenuate any tendency to perseverate on a particular lateral position but require greater attentional mobility than the procedure used in Experiment 1. Consequently, although 3 ½ month olds are capable of intermodal matching under this procedure (see Bahrick, 1988), it is possible that the procedure slightly underestimates their abilities relative to the two-trial procedure used in Experiment 1.

What information could the infants have used for matching the same gender faces and voices by 6 months or earlier? Some multimodal events or properties of objects are not amodally specified and therefore detection of intermodal correspondences may require intermodal learning from the start. We propose that there are at least three types of relations characterizing such multimodal events and speculate about the information to which infants may have responded.

One such type of relation may be characterized as "arbitrary but natural." For example, the relation between the sight of a person's face along with the unique sound of his or her voice is a natural relation that can only be learned through experience. Infants by 3 ½ months have already learned the idiosyncratic relation between the sight of their own mother's face and the sound of her voice (Burnham, Earnshaw, & Olymbios, 1986; Spelke & Owsley, 1979). Another example may be the relation between the particular odor of a nursing mother and the sight of her face, as distinct from that of an unfamiliar nursing woman. Recognition of the mother's odor has been shown by infants as young as 2 weeks (Cernoch & Porter, 1985). These relations always occur together in the environment, but there is little redundancy between information to the two sense modalities. Nevertheless, there are natural constraints on the nature of the sounds produced by a given person; the mother's voice must be within a typical pitch range, it is coincident with the sight of her face, it shares a synchrony relation with the movements of her mouth, and so forth.

A second type of arbitrary relation may be characterized as "arbitrary and artificial." These relations do not predictably occur together in nature and few constraints are built into the relation. Some examples include the particular sound of a telephone, the color and shape of an object in relation to the fundamental frequency of its sound, and the temperature or taste of an object in relation to its appearance or sound. The sound made by a telephone could as easily be a buzz or a ring, intermittent or continuous, loud or soft. A red toy that

is hitting a surface could make a higher or lower frequency impact sound and could just as easily be yellow or blue. These arbitrary relations must be learned through experience in each sense modality with the particular object. Little research has focused on infants' detection of these artificial, arbitrary relations. Reardon and Bushnell (1988) reported that 7 month olds learned to select from a pair of distinctively colored cups the one that had contained a sweet substance. Bushnell failed, however, to find evidence for infants' matching of color with temperature in an earlier experiment (Bushnell, 1986; see also Reardon, 1987). Along similar lines, Bahrick (1989) found that although 3 ½ month olds detected a change in amodal audiovisual relations (temporal synchrony; temporal information specifying object composition), they did not detect a change in an arbitrary audiovisual relation (pitch-color/shape) of an object hitting a surface.

Finally, one can also identify natural, complex classes of multimodal relations that are typical in the environment and are partially but not uniquely specified by amodal invariants. These may be characterized as "typical" relations. For example, heavy objects often produce deeper and louder sounds on impact than do lighter objects, but there is some overlap depending on the composition of the impacting object and surface. Detection of these relations may require more complex perceptual learning of arbitrary and/or amodal relations.

In our experiments, infants may have learned the intermodal correspondences over time, treating same gender faces and voices as if the relations were entirely arbitrary. That is, they may have learned to associate and integrate modality-specific face–voice information in their own environments and generalized to the specific individuals used in our experiments. On the other hand, they could have detected any number of invariant intermodal relations that typically define gender categories in the environment. For example, men typically have larger features, a more protruberant Adam's apple, a larger face and chest, more facial hair, or a face of a rougher texture than women. This typically corresponds to a voice with a lower pitch, a rougher sound, and vowels of a different formant frequency relative to women. Men may make larger, stronger body and facial motions with similarly modulated pitch-range variations that are distinguishable from those produced by women. Culturally defined differences are also evident, including shorter, flatter hair styles, no cosmetics, less jewelry, and so on, though it is difficult to imagine any auditory correlates that could be directly perceived. Thus, by 6 months, infants could have detected some or many of these typical but often overlapping audio–visual relations that specify gender categories. Infants need not integrate and associate auditory and visual information to perform this task. Given 6 months of perceptual learning in an environment replete with examples of gender relations, infants may have become attuned to salient audio–visual invariants defining these overlapping categories. In particular, by differentiating increasingly more specific relations over time, infants may initially detect face–voice correspondences on the basis of amodal relations (e.g., synchrony, common

rhythm). This in turn may lead to abstraction of more detailed information about the nature of the synchronous face and voice (i.e., typical relations just described), and finally association of arbitrary relations such as the relation between cosmetics and jewelry with voices of a higher pitch. This perceptual learning may enable more efficient abstraction of the relevant, typical relations in new instances of gender categories. Nevertheless, infants, by 6 months, detected audio–visual relations not by detecting temporal relations such as face–voice synchrony, because these amodal invariants which typically guide intermodal exploration were eliminated as a basis for matching in these experiments. Rather, they must have detected these relations based on gender-specific information. They showed this ability by successfully matching faces and voices across three different, novel pairs of male/female speakers.

One approach for disentangling explanations for the basis of infants' performance in the present task might be to determine whether there are any inherent constraints operating on the acquisition of intermodal knowledge about gender (e.g., Bahrick, 1988). Specifically, one could provide such infants who initially show no evidence of matching with appropriately versus inappropriately matched faces and voices (male face–female voice, versus male face–male voice, etc.) and assess under what combinations intermodal learning occurred. This study provides a first step in this direction, documenting an age group where matching does not occur. Another approach might be to alter systematically typical versus arbitrary face–voice relations (e.g., size of facial features and pitch of voice vs. presence of jewelry/makeup and pitch of voice) and determine under which conditions the matching effect is disrupted. A third approach would be to manipulate exposure time to information in one modality alone and then test for detection of intermodal relations. In this case we would expect to find a pattern parallel to one found in studies using a crossmodal method: Increasing the time for familiarization given to infants in one modality may lead to greater differentiation in both modalities at test (see Rose & Ruff, 1987, for a review). If this were found for infants who initially show no matching effect without exposure, it would suggest that infants are detecting relations inherent to the objects and events (typical or amodal relations) rather than using their knowledge of more arbitrary relations obtained prior to the experiment. Further research along these lines using events characterized by typical intermodal relations may help us better uncover the nature and basis of the developmental pattern for perceiving amodal, typical, and arbitrary intermodal relations.

ACKNOWLEDGMENTS

This research was supported by a National Institute of Mental Health Grant 1-R03-MH41411 to Arlene S. Walker-Andrews and by a National Institute of Child Health and Human Development Grant 5-R23-HD18766, to Lorraine E.

Bahrick. A portion of these data was presented at the 1988 International Conference on Infant Studies, Washington, DC.

We thank Lisa DeGisi, Debbie Krohn, Ruth Goldston, Gloria Peruyera, Jeffrey Pickens, and Lenore Szuchman for their assistance in the collection and/or analysis of data.

REFERENCES

Bahrick, L. E. (1983). Infants' perception of substance and temporal synchrony. *Infant Behavior and Development, 6*, 429–450.

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. (1989, July). *Perceptual differentiation of audio–visual events by infants.* Paper presented at The Fifth International Conference on Event Perception and Action, Oxford, OH.

Bahrick, L. E., Walker, A. S., & Neisser, U. (1981). Selective looking by infants. *Cognitive Psychology, 13*, 377–390.

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

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

Bower, T. G. R. (1974). *Development in infancy.* San Francisco: Freeman.

Burnham, D., Earnshaw, L. J., & Olymbios, P. (1986, April). *Facilitation of mother/stranger face discrimination by voices.* Paper presented at the International Conference on Infant Studies, Los Angeles.

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

Cernoch, J. M., & Porter, R. H. (1985). Recognition of maternal axillary odors by infants. *Child Development, 56*, 1593–1598.

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

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

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

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

Fagan, J. F., III. (1979). The origins of facial pattern recognition. In M. H. Bornstein & W. Kessen (Eds.), *Psychological development from infancy: Image to intention* (pp. 83–113). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Fagan, J. F., III., & Singer, L. T. (1979). The role of simple feature differences in infant recognition of faces. *Infant Behavior and Development, 2*, 39–46.

Francis, P. L., & McCroy, G. (1983, April). *Infants' bimodal recognition of human stimulus configurations.* Paper presented at the Society for Research in Child Development, Detroit.

Gibson, E. J. (1983). Development of knowledge about intermodal unity: Two views. In L. S. Liben (Ed.), *Piaget and the foundations of knowledge* (pp. 19–41). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Gibson, J. J. (1986). *The ecological approach to visual perception.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. (Original work published 1979)

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

Leinbach, M. D., & Fagot, B. I. (1986, April). *Gender-schematic processing by one-year-olds: Categorical habituation to male and female faces.* Paper presented at the International Conference on Infant Studies, Los Angeles.

Lieberman, P., & Blumstein, S. E. (1988). *Speech physiology, speech perception, and acoustic phonetics.* Cambridge, England: Cambridge University Press.

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., & Horowitz, F. D. (1980, April). *Integration of auditory and visual cues in speaker classification by infants.* Paper presented at the International Conference on Infant Studies, New Haven, CT.

Miller, C. L., Younger, B. A., & Morse, P. A. (1982). The categorization of male and female voices in infancy. *Infant Behavior and Development, 5,* 143–149.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24,* 175–184.

Piaget, J. (1954). *The construction of reality in the child.* New York: Basic Books.

Reardon, P. (1987). *Formation of arbitrary pairing of color and temperatures in 11-month-old infants.* Unpublished master's thesis, Tufts University, Medford, MA.

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

Rose, S. A., & Ruff, H. A. (1987). Cross-modal abilities in human infants. In J. D. Osofsky (Ed.), *Handbook of infant development* (pp. 318–362). New York: Wiley.

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

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

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

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

Tartter, V. C. (1986). *Language processes.* New York: Holt, Rinehart & Winston.

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

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

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

Walker-Andrews, A. S., & Lennon, E. M. (1985). Auditory–visual perception of changing distance by human infants. *Child Development, 56,* 544–548.