Journal of Experimental Child Psychology **81,** 93–115 (2002)
doi:10.1006/jecp.2001.2644, available online at http://www.idealibrary.com on IDEAL®

Infants' Ability to Match Dynamic Phonetic and Gender Information in the Face and Voice

Michelle L. Patterson and Janet F. Werker

University of British Columbia, Vancouver, British Columbia, Canada

Six experiments tested young infants' sensitivity to vowel and gender information in dynamic faces and voices. Infants were presented with side-by-side displays of two faces articulating the vowels /a/ or /i/ in synchrony. The heard voice matched the gender of one face in some studies and the vowel of one face in other studies and, in some studies, vowel and gender were placed in conflict. Infants of age 4.5 months showed no evidence of matching face and voice on the basis of gender, but were able to ignore irrelevant gender information and match on the basis of the vowel. Robust evidence of the ability to match on the basis of gender was not evident until 8 months of age. This set of findings suggests that, when identical stimuli are used, gender matching does not emerge until a later age than does phonetic matching. Results are discussed in relation to key theories of intermodal development. © 2002 Elsevier Science

Key Words: intermodal perception; infants; preferential looking; face processing; speech perception; gender perception; inverted faces.

From the moment an infant is born, human faces and voices pervade perceptual experience. Intermodal relations between faces and voices are crucial for the acquisition of linguistic, social, and emotional skills. Although infants have to learn most intermodal relations, it has been shown that by 4 months of age infants can detect many *amodal invariants* even if they have had little or no experience with the relations in question (Walker-Andrews, 1994). Amodal invariants are perceptual cues that are tied to the structural properties of an action or event and are not specific to a particular sensory modality. One amodal invariant, for example, is rhythm which can be detected by listening to a sound or by watching the visible event. Young infants are remarkably adept at matching amodal audio-visual events based on temporal synchrony, duration, rate, and even affective information in the face and voice (see Kellman & Arterberry, 1998; Rose & Ruff, 1987, for reviews). Such findings

This research was supported by the Social Sciences and Humanities Research Council (SSHRC) in the form of postgraduate fellowships to M.L.P. and by the National Science and Engineering Research Council (NSERC) in the form of an operating grant (No. OGP0001103) to J.F.W.

Address correspondence and reprint requests to Michelle Patterson, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver BC, Canada V6T 1Z4. E-mail: mpatters@interchange.ubc.ca.

0022-0965/02 $35.00

have led those who work from a Gibsonian perspective to argue that the senses are unified at birth and that perceptual development is characterized by differentiation of increasingly finer aspects of stimulation (Gibson, 1969). On the other hand, not all multimodal relations are specified by amodal invariants, nor do all researchers agree that any multimodal relation can be detected without some learning. Proponents of the integration view (Birch & Lefford, 1963; Piaget, 1952) claim that intersensory perception develops gradually as the child organizes modality-specific actions into a coordinated representation of the world.

Previous studies in our lab (Patterson & Werker, 1999) with both female and male faces replicate and extend past findings that 4.5-month-old infants can match phonetic information in the face and voice (Kuhl & Meltzoff, 1982, 1984; MacKain, Studdert-Kennedy, Spieker, & Stern, 1983; Walton & Bower, 1993) showing that even very young infants can detect bimodal invariants in speech. These findings suggest that phonetic information in the lips and voice may be another instance of an amodal invariant for infants. It is not known how this special amodal status emerges. It could be explained by a specialized module for speech in the brain, as suggested by proponents of the motor theory of speech perception (Liberman & Mattingly, 1985); alternatively, it could result from a history of interaction between a genetically initiated neural substrate and invariantly occurring species-specific experience (e.g., human speech; Werker & Tees, 1992). This epigenetic view has been adopted by infant speech perception researchers as reflecting a period of innately guided learning (Jusczyk & Bertoncini, 1988).

Gender is another salient, socially important, and naturally occurring class of information to which infants typically have been exposed. Gender information in the face and voice has been described as a *natural and typical* multimodal relation (Walker-Andrews, 1994). Natural and typical relations frequently occur but are neither exclusively amodal nor exclusively arbitrary. They may comprise sets of overlapping relations, some inherent in audio-visual structure and some that require a period of learning. The intermodal matching of *natural and typical* relations is expected to emerge later than that of *amodal* relations due to the culturally specific information that must be learned.

Only one set of studies has examined infants' ability to match gender using dynamic faces and voices (but see Poulin-Dubois, Serbin, Kenyon, & Derbyshire, 1995; Spelke & Owsley, 1979, for studies showing gender matching at 9–12 months of age using static facial displays). Walker-Andrews, Bahrick, Raglioni, and Diaz (1991) showed infants of ages 4 and 6 months videos of a male and a female face speaking side-by-side with a single soundtrack that corresponded to the gender of one visual display but was synchronized with the articulations of both displays. This study, in contrast to the work with vowel matching, suggests that the ability to match dynamic faces and voices based on gender cues may not be evident until 6 months of age and may not be based exclusively on the detection of amodal relations.

The ability to match gender in face and voice may depend upon the kind of gender information to which the infant is exposed. Visually, gender can be differentiated on the basis of both physiognomic properties such as size of features, especially the nose (Chronicle et al., 1995), and skin texture (Brown & Perret, 1993), as well as culture-specific variations such as facial hair, hair length, jewelry, and make-up. In terms of voices, men's tend to be lower pitched, although there is overlap. In addition, vowels produced by a man, woman, or child are different in terms of formant frequency (Peterson & Barney, 1952). Such differences result primarily from the length of the vocal tract and the size of the vocal folds (Leiberman & Blumstein, 1988) and thus may have visible correlates. For example, the human male adult tends to be larger than the female, and he has a longer and bigger throat, a broader nose, a more protuberant Adam's apple, a larger face, and broader shoulders.

Infants are able to detect the physiognomic properties of gender at a young age. One could argue that in order to match gender in the face and voice, infants need to both discriminate these cues and categorize on their basis. Habituation and visual preference studies have shown that infants well under one year of age can discriminate and categorize male and female faces (Cornell, 1974; Leinbach & Fagot, 1993; Lewkowicz, 1996, 1998) and voices (Miller, 1983). Together, studies of discrimination and categorization suggest that infants are able to notice gender cues by 3 to 5 months of age, but may not be able to use both seen and heard cues to categorize based on gender until the second half of the first year of life.

Culture-specific cues to gender tend to vary more than physiognomic cues and thus may require a longer period of learning (Leinbach & Fagot, 1993). Gender-specific cultural conventions for hair length, hairstyle, clothing, and jewelry (e.g., ear and nose rings) vary tremendously across cultures and even change across time within a culture. It is worth noting that, in their study of gender matching, Walker-Andrews et al. (1991) asked models to recite a nursery rhyme. With its greater variability in content, a nursery rhyme likely contains culturally richer cues to gender than do the isolated vowels that have been used in phonetic matching studies. Furthermore, the models in Walker-Andrews et al.'s studies wore gender-typical clothing and the woman wore make-up. It could be that these more richly textured stimuli prompted infants to use their knowledge of learned gender relations rather than relying on the amodal, physiognomic cues to gender. To the extent that those gender attributes in the face and voice which reflect cultural conventions require a period of learning, it is possible that infants might not be able to use them to detect gender until a later age than they would physiognomic relations.

Given the biological and social importance of gender, it is possible that gender attributes in the face and voice that are primarily physiognomic could be seen as another set of *amodal* relations (like phonetic relations). Such gender cues may be detected by infants younger than 6 months of age, perhaps as young as 4 months, the age at which infants detect most other amodal relations (see Kellman & Arterberry, 1998). Perhaps if the same simple dynamic stimuli were used for gender matching as have been used in phonetic matching tasks, infants would be better able to attend to the physiognomic aspects of the stimuli. This might enable gender matching at the same age as has been shown for vowel matching.

In order to directly test the question of whether infants can match gender information in the face and voice at the same age they can match phonetic information, we used the same stimuli as in Patterson and Werker's (1999) vowel matching studies. By presenting infants with isolated repeated vowels we hoped to make the attributes of the voice that are more reliant on physical differences between male and female vocal tracts more salient and deemphasize the cultural information that might be more easily displayed in recitations of nursery rhymes. Similarly, by having the male and female wear the same gender-neutral clothing and by making salient culture-specific visual information (i.e., make-up, jewelry) unavailable, we hoped to draw attention to the physiognomic features of the face. Using these same stimuli that had been used to assess vowel matching, we would be able to directly compare infants' ability to match gender information in face and voice to their ability to match phonetic information.

Six experiments are reported. In Experiment 1, we examine infants' ability to match face and voice on the basis of gender using the same stimuli as have been previously used to show vowel matching at 4.5 months of age. In Experiments 2–4, gender and phonetic information are placed in conflict to test their relative importance in bimodal matching in infants of age 4.5 months. Experiments 5 and 6 determine at what age infants match gender information in the face and voice with our stimuli. Together, these six experiments help us understand more completely the relative importance for infants of different kinds of dynamic information in the face and voice.

EXPERIMENT 1: GENDER MATCHING IN 4.5-MONTH-OLD INFANTS

The purpose of Experiment 1 was to determine whether 4.5-month-olds can match gender information across dynamic displays of the same faces and voices used in the previous vowel matching study (Patterson & Werker, 1999). If 4.5-month-olds can detect concomitant gender information in face and voice, they should look longer at the face that matches the gender of the heard sound. If gender matching, like phonetic matching, is possible with such carefully matched stimuli, this might suggest that gender cues are detected as an amodal relation. If infants do not match based on gender using our stimuli, this would support the suggestion that gender matching involves a combination of natural and typical relations and, thus, requires a longer period of learning than does phonetic matching.

Method

Participants

Mothers were recruited from a local maternity hospital shortly after giving birth or they responded to an advertisement in the local media. The final sample consisted of 64 infants, 32 male and 32 female, ranging in age from 19.1 to 23.2 weeks (M = 21.1 weeks, SD = 2.1 weeks). An additional 31 infants were tested and excluded from analyses due to fussiness (9), not looking at both stimuli during Familiarization (7), total looking time less than 1 min (7), looking at the same screen for the entire Test phase (2), equipment failure (4), and mother interference (2). Infants had no known visual or auditory abnormalities, including recent ear infections. Infants who were at-risk for developmental delay or disability (e.g., preterm, low birth weight) were not tested.

Stimuli

The same stimuli were used as in Patterson and Werker (1999). Multimedia computer software (mTropolis Version 1.1) on a Macintosh 7300 was used to combine, control, and present digitized audio and visual stimuli. Infants were shown two filmed images displayed on separate side-by-side computer monitors of a female and a male face articulating the same vowel (either /a/ or /i/) in synchrony. The sound track corresponding to the articulated vowel was presented through a speaker (Sony SRS-A60) midway between the two images. Since infants can detect face–voice correspondences based on temporal cues, the two visual images were presented in synchrony and the sound was aligned with the images so that it was equally synchronous with the onset of both mouth movements.

Female and male faces were selected for similar coloring (Caucasian, fair hair) and attractiveness. The female had shoulder-length hair and the male's hair was all-one-length extending below his ears. Both the female and male were filmed against a black background, both wore white turtlenecks, and neither wore jewelry or make-up. First, the male was filmed producing the vowel /a/ to the beat of a metronome set at 1 beat per 3 s. This 2-min recording was then played back over a TelePrompter and all other vowels (male /i/ and female /a, i/) were produced in synchrony with the male's /a/ (see Fig. 1).

As in Kuhl and Meltzoff (1984) and Patterson and Werker (1999), a different male and female were selected to record the audio stimuli. Different voices were used to ensure that there were no idiosyncratic cues linking a specific voice to a specific face. Audio recordings were made in a sound-proof recording booth using a studio-quality microphone and were recorded onto audio tape. Speakers were asked to articulate the vowels /i/ and /a/ with equal intensity and duration. The fundamental frequency at the maximum, minimum, and midpoint of each vowel for the male and female speaker was measured. As seen in Table 1, although intrinsic pitch made /i/ higher than /a/ for both the male and the female, the female voice is higher in pitch than the male voice at all times.

One visual /a/, one visual /i/, and one instance of each vowel sound for both female and male stimuli were chosen by three judges who rated what they deemed to be the five best visual and audio stimuli.[1] The facial images were cho-

[1]Kuhl and Meltzoff (1982) chose 20 audio and visual /a/s and /i/s to make two film loops. We digitized the audio and visual stimuli onto an I-Omega CD, thus the file size was limited such that we could pick only three instances of each audio and visual stimulus. When transferring these files to the multimedia authoring program, we chose one instance of each audio and visual stimulus to speed up running of the program.

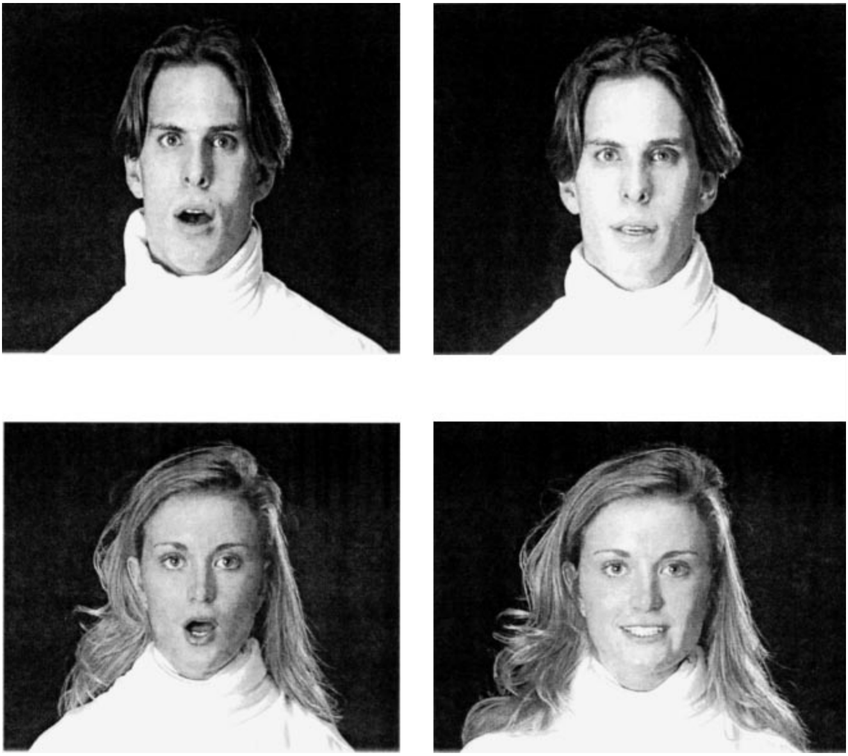

FIG. 1. Visual stimuli used in Experiments 1–3.

sen such that duration of individual articulations fell within a narrow range that overlapped for the two vowels, the head did not move, and one eye blink occurred after each articulation. For the female, the length of time that the lips were parted was 0.94 s for /a/ and 0.95 s for /i/. For the male, this duration was 1.27 s for /a/ and 1.28 s for /i/. A comparable process was used to select the audio stimuli. Since duration of mouth opening can be longer than sound duration but not vice

TABLE 1
Frequency (Hz) of Vowels Produced by Female and Male Speaker

	Maximum	Minimum	Midpoint
Female			
/a/	243.45	201.52	229.8
/i/	292.86	182.88	240.4
Male			
/a/	151.96	87.39	119.2
/i/	165.81	99.92	136.7

versa, we ensured that the vowel sounds were of the same or shorter duration than the mouth opening. For the female, duration of the sound was 0.61 s for /a/ and 0.63 s for /i/. For the male, this duration was 0.62 s for /a/ and 0.73 s for /i/.

The films and audio files were digitized and entered into a customized computer program (mTropolis, Version 1.1) which locked the faces in phase. Next, the sound was carefully synchronized with each visual stimulus so that it began 1 s (15 frames) after the mouth first started to move. Each articulation was repeated to form a continuous series of articulations occurring once every 3 s. When displayed on the monitors the faces were approximately life-size, 17 cm long and 12 cm wide, and their centers were separated by 41 cm. The sounds were presented at an average intensity of 60 $\pm$ 5 dB SPL.

Equipment and Test Apparatus

The stimuli were presented on two 17″ color monitors (Acana 17P) in the testing room. Black curtains covered the wall so that only the monitor screens and the camera lens (JVC GS-CD1U), positioned between and above the two monitors, were visible. The infant was seated 46 cm from the facial displays (visual angle subtended 29 degrees) in an infant seat secured to a table and the caregiver was seated behind the infant. The speaker was centered midway between the two monitors behind the curtain. During testing, a 60-W light in a lamp shade was suspended 110 cm above the infant.

Procedure

The experimental procedure involved two phases: Familiarization and Test. During the Familiarization phase, the articulating faces were presented without sound for 27 s. First, each visual stimulus (the articulating male and female face) was presented alone, one on each monitor, for 9 s each. During the final 9 s of the Familiarization phase, both articulating faces were presented simultaneously without sound.[2] Both stimuli were then occluded for 3 s before the Test phase began. During the 2-min Test phase, both faces were presented simultaneously, one on each monitor, and one of the two sounds (either /a/ or /i/) was played repeatedly. Thus, each infant heard only one sound. Sound presented, left–right positioning of the two faces, order of familiarization (right first vs left first), and infant sex were counterbalanced across infants.

Scoring

Coding was performed using a Panasonic video recorder which allowed frame-by-frame analysis. Coders were undergraduate students who were blind to the stimuli presented to the infant. Interobserver reliability was assessed by

[2]Kuhl and Meltzoff (1982) did not include the simultaneous presentation of both faces in their Familiarization phase. This phase is typically included in studies of infant word comprehension (e.g., Hirsch-Pasek & Golinkoff, 1992). The logic behind including this phase is to teach infants that both displays can be on simultaneously and it can be used as a check for infant side bias.

rescoring 25% of the participants. Duration of gaze was scored for each second when the infant was looking either to the right or to the left monitor. Frame-by-frame coding was used to determine if the point of gaze-shift occurred before or after the midpoint of the second. If the infant fixated the screen for more than half the second, the infant was coded as looking at the screen for the whole second. If the infant looked less than half a second, the infant was coded as not looking during that second. Individual gaze-on seconds were summed for each display and divided by the total time spent looking at the displays to obtain the percentage of total looking time (PTLT) spent on each display during the Test phase as well as for the 9-s period of the Familiarization phase where both faces were presented simultaneously. PTLT to the match was also calculated for the first and second minutes of the Test phase separately. Finally, the longest look to the match and the mismatch was recorded for each infant and summed across infants. The percentage agreement for each second in the sampled periods ranged from 95.9 to 99.4% (M = 97.9%) for infant looking.

Results and Discussion

Infants spent 74.2% of the total Test phase looking at either of the two faces. According to a paired t test, no side bias was evident during the Familiarization phase ($p > .05$). Overall, infants of age 4.5 months showed no evidence of matching on the basis of gender. On average, infants spent 52.4% of the total looking time on the gender match, which was not significantly greater than chance (50%; $p > .05$). Of the 64 infants tested, only 38 looked longer at the match than the mismatch (binomial test, $p > .05$). Using PTLT to the match as the dependent variable, an omnibus ANOVA was conducted to explore any possible interactions among four primary variables (Sex of infant, Side of match, Heard vowel, and Gender of face/voice). All main effects and interactions were nonsignificant. Other factors counterbalanced in the design were entered into separate one-way ANOVAs; there were no significant differences in looking time to the match based on order of familiarization or side of gender.

When infant looking was examined during the first and second minutes of the Test phase there was still no evidence of matching based on gender information [PTLT(min1) = 51.8; PTLT(min 2) = 55.0, $p > .05$]. Similarly, infants' longest looks to the match (19.31 s) versus the mismatch (19.56 s) were very similar. Therefore, the 4.5-month-olds tested in this study, an age at which infants are able to match phonetic information in the face and voice, were not able to match gender information using the same stimuli.

Research has shown that infants between 4 and 6 months of age are highly sensitive to and interested in vowel information (Kuhl et al., 1992; Polka & Werker, 1994). It is possible that the infants in Experiment 1 failed to match on the basis of gender because they attempted instead to match the vowel information in face and voice. That is, vowel information may have been more salient for infants than gender information. Since the heard vowel matched both faces in terms of phonetic information, infants may have stopped looking for a "match." In our next experiment we tested the plausibility of this hypothesis by placing vowel information in conflict to see if infants match on the basis of gender without the potentially distracting presence of a vowel match.

EXPERIMENT 2: GENDER MATCHING WITH DISCREPANT VOWEL

If 4.5-month-old infants can match gender information in the face and voice but were distracted in Experiment 1 by the presence of matching phonetic information, then infants may look preferentially to the gender match if there is no matching phonetic information. To test this possibility, infants were presented with a male and a female face, each articulating the same vowel sound (either /a/ or /i/). One of the faces matched the gender of the voice and the other did not. Of critical importance, the heard vowel did not match the articulated vowel on either of the two faces. For example, an infant might be shown displays of a male and a female face each articulating the vowel /a/. They would then hear a female voice saying /i/. By removing the possibility of a phonetic match, we hoped to more unambiguously focus the infants' attention on the gender of the face and voice.

Method

Participants

The final sample consisted of 32 infants, 16 male and 16 female, ranging from 17.5 to 23.1 weeks (M = 19.4 weeks, SD = 1.9 weeks). An additional 15 infants were tested but excluded from analyses due to fussiness (5), failing to look at both screens during Familiarization (4), looking at the same screen during entire Test phase (2), equipment failure (1), and mother interference (3). Equipment, test apparatus, and scoring were identical to those in Experiment 1.

Stimuli and Procedure

As in Experiment 1, infants were seated in front of two side-by-side visual displays, one presenting a female face and the other presenting a male face, both articulating the same vowel sound (either /a/ or /i/) in synchrony. However, unlike Experiment 1, the heard vowel did not match the articulating faces in terms of vowel. All other aspects of the stimuli and procedure were identical to those of Experiment 1.

Results and Discussion

Again, infants showed no evidence of side bias during the Familiarization phase (paired t test, $p > .05$). Infants spent 84.2% of the Test phase fixating one of the two faces. As illustrated in Fig. 2, infants did not show a visual preference for the face that matched the gender of the voice despite nonmatching phonetic information (PTLT = 56.28; $p > .05$). A four-way ANOVA (Infant sex, Side of match, Heard vowel, Gender of voice) revealed no significant main effects or interactions. The effect was similar across the first minute (PTLT = 56.11) and

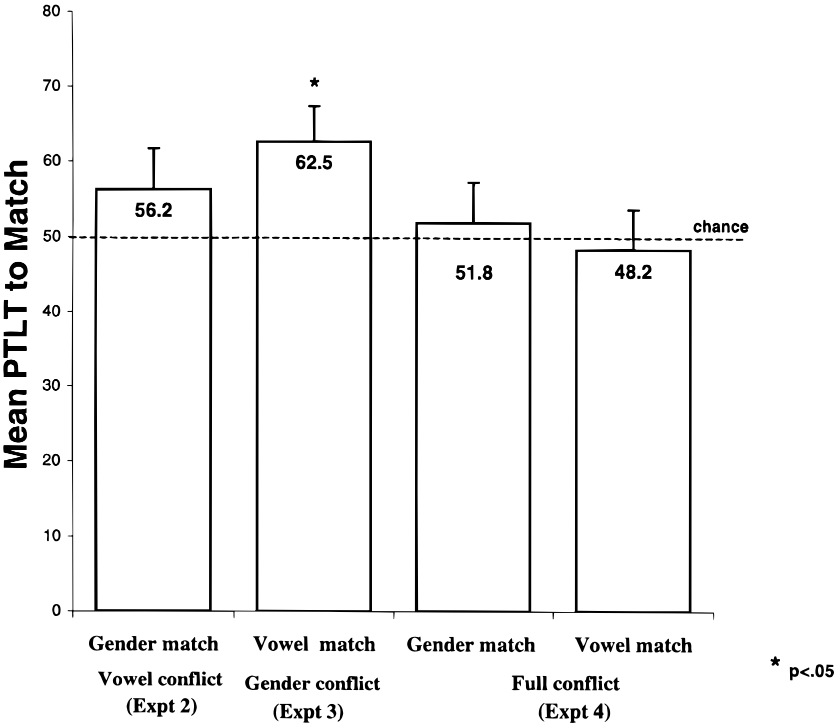

FIG. 2. Mean percentage of total looking time (PTLT) to the face that matched the heard voice (Experiments 2 and 3) and to both the gender and the vowel match (Experiment 4). Error bars represent standard error.

the second minute (PTLT = 56.61) of the Test phase and only 19 of 32 infants looked longer at the gender match (binomial test, $p > .05$). Finally, infants' longest look to the match (25.24 s) versus the mismatch (19.28 s), although suggestive, did not differ significantly according to a paired t test ($p > .05$).

In Experiment 2, we attempted to remove vowel as a source of matching by presenting different phonetic information in face and voice. Infants of age 4.5 months still showed no preference for the gender match. Both PTLT and longest look to the match were marginally longer than in Experiment 1, but the effect was not significant. This finding suggests that, even when a potentially distracting phonetic match is not possible, infants still do not match gender cues in face and voice. In reminder, earlier work (Patterson & Werker, 1999) has shown that 4.5-month-old infants were able to match vowel cues in the face and voice when the gender cues were congruent. The results of Experiments 1 and 2, in comparison to the earlier work, show that infants do match phonetic information before they are able to match gender cues in face and voice. Moreover, infants' failure to match on the basis of gender, even when vowel information is removed, raises the possibility that they may not even detect the gender-relevant information at this age. In the next experiment, we attempt to understand this further by seeing whether 4.5-month-olds match phonetic information in the face and voice when the gender information in the faces is incongruent with the heard voice, or whether the presence of conflicting gender information disrupts vowel matching.

EXPERIMENT 3: VOWEL MATCHING WITH DISCREPANT GENDER

The purpose of Experiment 3 was to see if infants still match faces and voices based on vowel information when the gender of the voice does not match the gender of either face. Infants of age 4.5 months do match phonetic information when the face and voice are of the same gender (Patterson & Werker, 1999). Moreover, bimodal speech perception in adults is unaffected by mismatching gender. Green and Kuhl (1991) presented adults with stimuli wherein a male voice saying /ba/ had been dubbed onto a female face articulating /ga/ and the subject perceived "da" as in the classic "McGurk" finding (McGurk & MacDonald, 1976). Conflicting gender information in this case had no significant impact on the McGurk effect in adults even though the discrepancy was easily detected by the subjects. If infants are like adults, they may match phonetic information even when the gender information is incongruent.

Method

Participants

The final sample consisted of 32 infants, 16 male and 16 female, ages 16.2 to 22.3 weeks (M = 19.8 weeks, SD = 1.4 weeks). An additional 12 infants were tested but excluded from analyses due to fussiness (5), not looking at both screens during familiarization (3), looking at the same screen for entire Test phase (1), mother interference (2), and equipment failure (1). Stimuli, test apparatus, and scoring were identical to those of Experiment 1.

Procedure

The procedure was identical to that of Experiment 1 except that the soundtrack corresponded to the vowel articulation of one of the faces but was of the opposite gender. Therefore, infants were presented with either side-by-side male or female faces, each articulating a different vowel sound (either /a/ or /i/) in synchrony. Gender of the heard sound did not match either of the seen faces; however, the heard vowel matched one of the two articulating faces. Thus, the heard sound matched one of the articulating faces in terms of phonetic information but neither face in terms of gender. For example, an infant might see two female faces, one articulating the vowel /i/ and one articulating the vowel /a/, but would hear a male voice saying /a/. If they do match vowels even with conflicting gender in this condition, they should look longer to the face articulating /a/.

Results and Discussion

No evidence of side bias was observed during the Familiarization phase (paired t test, $p > .05$). Infants spent 78.3% of the total Test phase fixating one of the two faces. As illustrated in Fig. 2, infants spent significantly more time looking at the face that matched the heard vowel regardless of the discrepant gender (PTLT = 62.2; $t(31) = 2.39, p < .05$). Similarly, the effect was present in both the first minute (PTLT = 61.0; $t(31) = 1.99, p < .05$) and second minute (PTLT = 64.5; $t(31) = 2.91, p < .05$) of the Test phase and 24 of 32 infants looked longer at the sound-specified face (binomial test, $p < .05$). Infants' longest looks during the Test phase were significantly longer to the match (32.4 s) versus the mismatch (13.6 s) according to a paired t test ($t(31) = 2.61, p < .05$).

A four-way ANOVA (Infant sex, Side of match, Heard vowel, Gender of voice) revealed significant main effects for Side of match ($F(1, 16) = 25.2, p < .01$) and Heard vowel ($F(1, 16) = 19.8, p < .05$). Overall, infants looked longer at the match when it was on the right side (PTLT = 75.0) versus the left side (PTLT = 48.5) and when the heard vowel was /a/ (PTLT = 73.7) versus /i/ (PTLT = 51.5). There was also a significant Side of match × Heard vowel interaction ($F(1, 16) = 13.2, p < .01$); infants looked longer at the match when it was on the left-hand side if the heard vowel was /a/ rather than /i/. A significant Side of match × Gender of voice interaction ($F(1, 16) = 4.75, p < .05$) revealed that infants looked longer at the match when it was on the left-hand side if the gender of the heard voice was female rather than male. Finally, there was a significant three-way interaction among Infant sex × Side of match × Gender of voice ($F(1, 16) = 7.62, p < .05$); boys looked longer at the match on the right-hand side when the female voice was played and longer at the match on the left-hand side when the male voice was played. Girls' looking time to the match on the left and right side was not influenced by gender of heard voice. Due to the significant right-side bias, we applied the correction for side bias described by Humphrey and Tees (1980) to the data. The corrected proportions did not change the results reported using uncorrected looking times; there was still a significant overall preference for the match ($p < .05$).

These results, like those reported with adults (Green & Kuhl, 1991), suggest that 4.5-month-old infants are able to ignore incongruent gender information in order to match on the basis of phonetic information in face and voice. This stands in contrast to the results from Experiment 2 in which infants of this age were not able to match based on gender despite discrepant phonetic information. Thus, Experiment 3 provides additional evidence that phonetic information is particularly salient for infants at 4.5 months of age. It should be noted, however, that while the majority of infants preferred to look at the face that matched the heard sound in terms of vowel and the PTLT (62.2) was very close to that reported in Patterson and Werker (1999) for phonetic matching with congruent gender (62.7 for male stimuli and 64.7 for female stimuli), there was considerably more variability in Experiment 3 than in the Patterson and Werker (1999) vowel matching study. In that study, there was no significant side bias or vowel effect, nor were there any significant interactions. The side bias, vowel effect, and interactions reported in Experiment 3 reveal that some infants may have been confused by the mismatching gender information and showed only side preferences[3] or other idiosyncratic behaviors. This raises the possibility that gender information may be noticed, to some extent, by infants at 4.5 months of age, even if they cannot yet use gender information to form an integrated percept.

EXPERIMENT 4: VOWEL AND GENDER INFORMATION IN CONFLICT

If infants are born with a unity of the senses and expect certain sounds to be concomitant with certain events (Morrongiello, 1994) then conflicting audiovisual information, particularly when synchrony is maintained, might be especially distressing or attention-getting. If infants do not detect gender cues in our stimuli, they should not provide a source of conflict. However, if, as hinted by Experiment 3, gender is detected, it could interfere with vowel matching. To provide the strongest possible test of this hypothesis, in Experiment 4 we placed gender *and* vowel information in full conflict to see if infants show any matching—be it gender or phonetic—under these conditions. For example, an infant might *see* a male face articulating /a/ and a female face articulating /i/ and *hear* a male voice saying /i/. Therefore, infants could look at the face that matched the heard sound in terms of vowel, or in terms of speaker gender, or neither.

If 4.5-month-olds do not match gender information in the face and voice with our stimuli because they do not perceive gender cues as such then, when vowel and gender cues are placed in direct conflict, mismatching gender should not interfere with infants' ability to match based on vowel information. However, as suggested by Experiment 3, if infants, at some level, detect equivalent gender information in face and voice, one might expect this information to interfere with their ability to match based on vowel information. To the extent that conflicting audiovisual information is disruptive, if gender information is detected then it should interfere maximally with infants' ability to match on the basis of phonetic information in this full conflict manipulation.

Method

Participants

Recruitment procedures were identical to those of Experiment 1. The final sample consisted of 32 infants, 16 male and 16 female, ranging in age from 17.9 to 22.2 weeks (M = 19.3 weeks, SD = 1.3 weeks). An additional 13 infants were tested but excluded from analyses due to fussiness (4), not looking at both stimuli during Familiarization (3), total looking time less than 1 min (3), and looking at the same screen for the entire Test phase (3).

[3]Due to the significant right-side bias, we applied Humphrey and Tees's (1980) correction procedure to the data from Experiment 3. Even with the correction for side bias, PTLTs to the match continued to be significant when the match was on the left side (.62) as well as the right side (.63).

Stimuli and Procedure

All aspects of the stimuli, test apparatus, and procedure were identical to Experiment 1 with the following exceptions. Infants were shown side-by-side images of a male and a female face on separate monitors, each articulating a *different* vowel (either /a/ or /i/). The sound track corresponded to the gender of one of the faces but matched the other face in terms of vowel articulation. The two visual images were temporally synchronous with the vowel sounds.

Results and Discussion

No side bias was evident during the Familiarization phase (paired t test, $p >$ .05). Overall, infants spent 87.5% of the total Test phase fixating one of the two faces. During the Test phase, infants spent approximately equal amounts of time fixating the gender match (PTLT = 51.8) and the vowel match (PTLT = 48.2) according to a paired t test; thus, infants did not show a visual preference for either gender or linguistic information (see Fig. 2). Of 32 infants, 16 looked longer at the gender match while 16 looked longer at the vowel match.

All statistics were conducted using PTLT to the vowel match as the dependent variable. A four-way ANOVA (Side of vowel match, Infant sex, Gender of voice, Heard vowel) revealed no significant main effects for Infant sex, Gender of voice, or Heard vowel; however, there was a significant main effect for Side of vowel match ($F(1, 16) = 7.37, p < .05$) and its interaction with two other variables, Infant sex ($F(1, 16) = 6.57, p < .05$) and Heard vowel ($F(1, 16) = 5.98, p < .05$). Boys, but not girls, looked longer at the vowel match when it was on the right-hand side rather than on the left-hand side and, overall, infants looked longer at the vowel match when it was on the right-hand side if the heard vowel was /i/ rather than /a/. Again, due to the significant right-side bias, we applied the correction for side bias described by Humphrey and Tees (1980) to the boys' data. The corrected proportions did not change the results reported using uncorrected looking times. Paired t tests revealed that PTLT spent on the vowel match did not significantly differ from the first minute (45.0) to the second minute (53.7) nor did infants' longest look to the vowel match (21.8 s) differ from longest look to the gender match (21.2 s, $p > .05$).

It seems that simultaneously conflicting vowel and gender information entirely disrupted infants' ability to match on the basis of phonetic information. This finding is consistent with the hypothesis that 4.5-month-olds do notice gender-relevant cues but are unable to use the auditory specification of gender to guide visual exploration of facial cues when both gender and phonetic information are varying in the face. Several researchers (see Bahrick & Pickens, 1994) have described a lag between infants' noticing information and their ability to use auditory information to guide visual exploration of that information when irrelevant visual information is simultaneously present. The results from Experiments 3 and 4 provide support for this hypothesis.

If the hypothesis that 4.5-month-old infants are at least noticing mismatching gender is correct, and if detection precedes more functional use, one would expect to see evidence of gender matching emerging at a slightly older age. The next experiment was designed to test this hypothesis with slightly older infants.

EXPERIMENT 5

The only previous study examining infants' ability to match dynamic audio and visual information based on gender cues reported evidence of matching at 6 months of age (Walker-Andrews et al., 1991). The results reported in the current set of studies suggest that infants at 4.5 months of age detect gender information in face and voice but are not yet able to use it to guide their looking preferences. Thus, by 6 months of age, infants may be able to match on the basis of gender using our stimuli. Experiment 5 was designed to specifically test this hypothesis with 6-month-olds in the same task used in Experiment 1.

Method

Participants

Recruitment procedures were identical to Experiment 1. The final sample consisted of 32 babies ranging in age from 24.9 to 27.1 weeks ($M = 26.6$ weeks, $SD = 1.9$ weeks). An additional 13 infants were tested but excluded from analyses due to fussiness (7), not looking at both stimuli during Familiarization (4), and equipment failure (2). Equipment, procedure, and scoring were identical to those of Experiment 1.

Results and Discussion

A paired *t* test indicated that infant looking during the Familiarization phase was not biased to either the right or the left side. Overall, 6-month-olds spent 77.4% of the total Test phase fixating one of the two faces. The ability of 6-month-old infants to match faces and voices based on gender appeared to be stronger than at 4.5 months but still not significantly above chance and not as robust as vowel matching at 4.5 months. Overall, 6-month-olds spent 58.1% of the total looking time on the match ($t(31) = 1.66, p = .11$; see Fig. 3) and of the 32 infants tested 22 looked longer at the match (binomial test, $p > .05$).

A paired-sample *t* test indicated that looking time to the match was similar across infants during the first minute (57.0%) and second minute (58.3%) of the Test phase ($p > .05$). Overall, paired-sample *t* tests indicated no significant difference between the longest look to the match (16.9 s) versus the mismatch (11.8 s) during the Test phase ($p > .05$); however, the duration of longest look for girls was significantly longer to the match (18.9 s) versus the mismatch (9.2 s, $t(15) = 2.16, p < .05$). Several studies (e.g., McClure, 2000) that have examined infant perceptual development have found that girls are often precocious in their abilities relative to boys. The findings from Experiment 5 are consistent with this pattern. Often, when an ability is present in female infants at one age, it appears in males a few months later (e.g., Hall, Belanger, & Lee, 2000).

A four-way ANOVA (Infant sex, Side of match, Heard vowel, Gender of voice) revealed a significant main effect for Side of match ($F(1, 16) = 8.05, p < .01$);

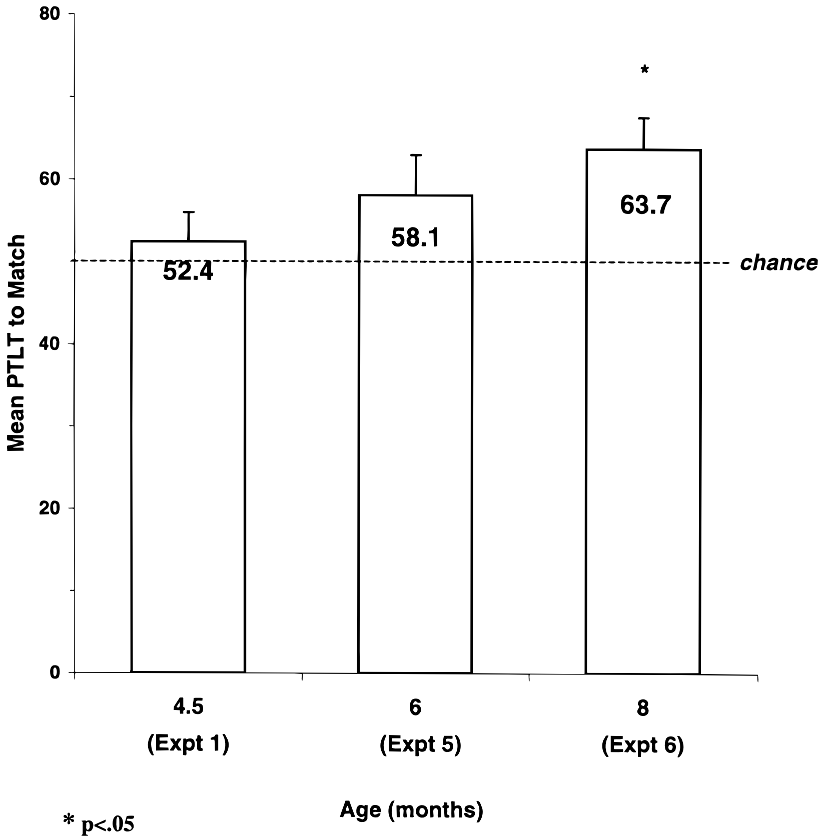

FIG. 3. Mean percentage of total looking time (PTLT) to the face that matched the gender of the heard sound as a function of infants' age in months. Error bars represent standard error.

infants looked longer at the match when it was on the right side (PTLT = 70.4) versus the left side (PTLT = 45.7). After applying Humphrey and Tees's (1980) correction for side bias, the effect was still nonsignificant. All other main effects and interactions were nonsignificant. Separate one-way ANOVAs revealed no preference for either the female or the male face or voice nor was there any significant effect for side of gender or order of familiarization. One of the dependent variables did suggest, however, that 6-month-old girls may be able to use knowledge of gender information in our stimuli to guide looking preferences. The final study was designed to determine exactly when infants are able to match gender information in dynamic faces and voices.

EXPERIMENT 6

Although 6-month-old girls showed an emerging preference for the match in longest look to the match versus mismatch but not in overall PTLT to the match.

There was no evidence that boys of age 6 months match faces and voices based on gender. It appears that, overall, 4.5- and 6-month-old infants are unable to match gender with our brief auditory stimuli and when typical cultural cues to visual gender are minimized. By 8 months of age, infants have had more experience with different-gender faces and voices. Thus, if the data from the girls at 6 months did indeed indicate an emerging preference, we expected to see a robust preference for the gender match in girls by 8 months and possibly in boys as well.

Method

Participants

Recruitment procedures were identical to those of Experiment 1. The final sample consisted of 32 infants, ranging in age from 33.6 to 38.3 weeks ($M = 36.7$ weeks, $SD = 2.1$ weeks). An additional 19 infants were excluded from analyses due to fussiness (8), not looking at both stimuli during Familiarization (4), total looking time less than 1 min (5), and locking onto one screen for the entire Test phase (2). Five of these infants had participated in the gender matching task at 4.5 months of age. All aspects of the stimuli, equipment, procedure, and scoring were identical to those of Experiment 1.

Results and Discussion

Overall, 8-month-olds spent 79.1% of the Test phase fixating one of the two faces. Unlike the younger infants, 8-month-olds looked significantly longer at the face that matched the gender of the heard voice. As shown in Fig. 3, overall infants spent 63.7% of the total looking time on the sound-specified face ($t(31) = 3.44, p < .01$). Of the 32 infants tested, 26 looked longer at the match than at the mismatch (binomial test, $p < .05$). The effect was also present in both the first minute (PTLT = 62.5; $t(31) = 3.75, p < .01$) and the second minute (PTLT = 67.2; $t(31) = 4.49, p < .01$) of the Test phase and infants' longest looks during the Test phase were significantly longer to the match (17.7 s) versus the mismatch (8.1 s) according to a paired t test ($t(31) = 2.84, p < .01$).

A four-way ANOVA (Side of match, Infant sex, Heard vowel, Gender of voice) revealed a significant main effect for Side of match ($F(1, 16) = 8.35, p < .01$); both sexes looked longer at the match when it was on the right-hand side (PTLT = 74.5) versus the left-hand side (PTLT = 53.5). The effect was still significant after we applied Humphrey and Tees's (1980) correction for side bias to the data. All other main effects and interactions were nonsignificant. Separate one-way ANOVAs revealed no significant differences in looking time based on order of familiarization, side of gender, and gender of voice or face ($p > .05$). An analysis omitting the five infants who had participated at 4.5 months revealed the same pattern of results.

The results above suggest that by 8 months of age infants show robust performance on a gender matching task. We conducted one further set of analyses to compare the PTLT to the gender match at the three ages tested in Experiments 1, 5, and 6. A priori comparisons of PTLTs among the three age groups tested in the gender matching task were carried out by performing three independent-sample t tests. The PTLTs for the 4.5- and 6-month-olds did not differ significantly ($t(94) = .776, p > .05$) nor did the PTLTs for the 6- and 8-month-olds ($t(62) = .924, p > .05$). However, the PTLT for the 8-month-olds was significantly different than that for the 4.5-month-olds ($t(94) = 1.815, p < .05$). A trend analysis suggested a linear trend in the PTLT across Experiments 1, 5, and 6 ($F(1, 125) = 3.01, p = .06$), indicating that performance at 6 months of age was intermediate between that at 4.5 and 8 months of age. The mean PTLT to the match across the three age groups is summarized in Fig. 3. Thus, by 8 months of age, infants looked significantly longer at the face specified by the gender of the voice than they did at 4.5 months of age, with performance at 6 months intermediate between the two.

GENERAL DISCUSSION

Previous research has shown that infants as young as 4.5 months of age match acoustically presented vowel sounds with the appropriate facial articulation (Kuhl & Meltzoff, 1982, 1984; Patterson & Werker, 1999). These findings support claims that young infants can perceive structural correspondences between audio and visual aspects of speech input and are consistent with the possibility that phonetic information is represented amodally. It remained unclear, however, whether the intermodal perception of other biologically significant face–voice events shows the same early emergence. We chose to study gender because it is central to human social functioning. The current set of experiments were designed to explore the bimodal matching effect using both phonetic and gender information in dynamic faces and voices.

In Experiment 1, 4.5-month-olds showed no evidence of matching gender information in the face and voice. Although the heard vowel matched only one face in terms of gender, it matched both faces in terms of vowel. Since infants are interested in vowels at this age (Polka & Werker, 1994), they may have focused primarily on vowel information and may have neglected to look for a gender match. We tested this possibility in Experiment 2 by neutralizing the informational value of phonetic cues in the lips; however, infants still did not look preferentially at the gender match. Thus, it seems that 4.5-month-olds' failure to match based on gender was not entirely due to their matching phonetic information in the face and voice instead.

When one of the articulating faces clearly matched the heard vowel but the gender of the voice did not match either face (Experiment 3), infants looked significantly longer at the face that matched the heard vowel. This shows that 4.5-month-old infants are able to neutralize conflicting gender information and recognize the vowel match. However, the unexpected main effects and interactions raised the possibility of some interference from the mismatching gender cues. To explore this possibility, phonetic and gender information were placed in full conflict in Experiment 4. Here, 4.5-month-olds did not show a visual preference for either the phonetic or the gender match. The disruption of vowel matching caused by a gender mismatch suggests that infants do notice some of the bimodal information specifying gender; however, infants at 4.5 months of age may not yet be able to use this information to guide looking preferences in a gender matching task.

If infants detect but do not match gender information in face and voice at 4.5 months of age, when does the ability to match emerge? Experiment 5 was conducted to see if infants at 6 months match gender in our faces and voices. Like the 4.5-month-olds, 6-month-old infants' looking time to the match was not significantly greater than chance; however, evidence of preference was apparent in 6-month-old girls in one dependent variable, longest look. The final experiment tested an even older age group, 8-month-olds, in our gender matching task. The ability to match gender information in face and voice was unambiguously evident in both girls and boys at 8 months of age. It appears that infants of 4.5 months of age do notice the link between heard and seen gender since mismatching gender disrupts, partially (Exp. 3) or fully (Exp. 4), their ability to match on the basis of vowel. Nevertheless, infants at 4.5 months were not reliably able to match on the basis of gender using our stimuli (the same stimuli which enabled phonetic matching at 4.5 months) until 8 months of age.

It is of interest to compare our results with those previously reported in the literature. In reminder, previous reports using static faces and voices failed to find gender matching until 12 months of age. With dynamic faces, Walker-Andrews et al. (1991) failed to find robust gender matching at 4 months of age; however, matching was reported at 6 months of age. Although the results reported here are quite similar to those reported by Walker-Andrews et al. (1991), it is useful to consider why infants in our study did not show robust evidence of gender matching until a slightly later age than did infants in Walker-Andrews et al.'s study.

One major difference between our stimuli and those used by Walker-Andrews et al. (1991) concerned the audio stimuli. Walker-Andrews et al. recorded the models' voices separately from the visual stimuli; however, the voices did belong to the visual stimuli. Thus, there could have been an idiosyncratic relation between the matching face and voice (e.g., rising intonation paired with eyebrow raising) or a structurally specified relation between the physical features of the vocal organs (mouth and neck) and the sound of the voice. Our use of voices that did not belong to either of the faces along with the use of isolated vowels ensured that the faces and voices of both models were equally well-synchronized and that matching was not based on any idiosyncratic cues in the face and/or voice.

Although both our and Walker-Andrews et al.'s stimuli conveyed neutral facial and vocal affect, our stimuli were designed to minimize as many culture-specific cues as possible and to allow infants to focus on the structural features specifying gender in face and voice. Thus, we asked our models to remove all jewelry and make-up, the male's face was clean-shaven, and the speakers both wore white turtlenecks, which fully concealed any throat cues and removed any gender cues based on clothing. Remaining culture-specific cues were hair length (our female had long hair and our male had hair all-one-length just past his ears) and plucked eyebrows (female only). Thus, the visual stimuli eliminated many of the arbitrary face–voice associations related to gender, making it more likely that gender matching was based on physiognomic cues in the face (e.g., nose, jaw, cheekbones, skin texture).

The voice stimuli were also created in a way that minimized cultural display rules for gender. As noted above, we chose to use isolated vowels whereas Walker-Andrews et al. used nursery rhymes. The acoustic cues that specify that a sound has been produced by a shorter (female) or longer (male) vocal tract are contained even in isolated vowels. However, isolated vowels do not give the speaker the opportunity to introduce cultural specific display rules of speech as does continuous speech and nursery rhymes in particular. For this reason, we expected isolated vowels to bias infants to attend to structural cues rather than cultural display rules in the voice.

The highlighting of physiognomic cues and minimizing of culture-specific cues allowed us to assess whether gender matching, like phonetic matching, may be based at least initially on amodal properties in the face and voice. Rather than revealing an earlier emerging ability to match gender in the face and voice, however, infants presented with our stimuli demonstrated gender matching at a slightly later age than reported by Walker-Andrews et al. These findings show that matching of gender in the face and voice is not as immediately available to infants as is matching of phonetic and other amodal events. Furthermore, these findings are more consistent with the notion, suggested by Walker-Andrews (1994), that gender relations are not amodal but involve both typical and arbitrary information and thus may require a longer period of learning. Indeed, the slightly earlier age of matching seen in the Walker-Andrews et al. study may indicate that evidence of this learning is easier to see when there is more culturally rich information in the face and voice.

Bahrick and Pickens (1994; see also Walker-Andrews, 1994) suggest that discrimination of modality-specific cues, along with the detection of amodal invariants, may precede and guide learning about arbitrary object–sound relations. Such early perceptual abilities may direct infants' attention to appropriate object–sound pairings and then promote sustained attention and further differentiation. Initial detection of an amodal relation (e.g., voice–lip synchrony, shared rhythm, and tempo) may enable the infant to focus on a unitary event (e.g., the mother's face and voice). This, in turn, may lead to differentiation of more specific, arbitrarily paired audible and visible attributes (e.g., the sound of the voice with the unique appearance of the face). Bahrick (1992) reported a developmental progression across age where infants detected amodal relations at a younger age than arbitrary relations from the same events. Our results support this developmental pattern; that is, infants detect the amodal information specifying the phonetic match in the face and voice at an earlier age than they do the gender information in the same displays. They are, however, able to discriminate the gender of the face and voice alone in similar displays (e.g., Fagan, 1976; Miller, 1983). Perhaps, as suggested by Bahrick and Pickens (1994), the early attentional bias to some amodal properties, paired with discrimination of gender cues, sets the stage for increased learning about gender with development and experience.

An alternative explanation is that infants under 8 months of age have difficulty matching gender information in face and voice because of where young infants focus their attention in a visual display (see Lewkowicz, 1999). Object perception undergoes a major shift at about 7 months of age (Kellman & Arterberry, 1998). Before this time, visual processing is apparently based on an edge-insensitive process; therefore, 4.5- and 6-month-olds may not be as sensitive as older infants to the edge relationships that normally specify objects and events. Instead, they may respond primarily to kinematic properties, such as lip movements, thus enabling 4.5-month-olds to match faces and voices based on phonetic information. By 7 months of age, an edge-sensitive process takes over and provides the infant with a richer view of the world since kinematic properties as well as object connectedness and forms of hidden boundaries are specified. Thus, older infants can perceive objects and events in a more detailed and accurate fashion. This perceptual shift, along with the more general process of perceptual differentiation, may in part account for the age differences in infant responsiveness to bimodal gender information. It would be of interest to explore this explanation in future research.

In summary, the present studies with 4.5-, 6-, and 8-month-old infants demonstrate a developmental progression toward a robust ability to match faces and voices based on gender information. Using the same stimuli, infants of age 4.5 months were able to match phonetic information in face and voice; however, it was not until 8 months of age that infants were able the match the same faces and voices based on gender information. Our results also suggest that 4.5-month-olds notice gender-relevant cues but may be unable to use them to guide preferential looking when both vowel and gender information are varying in the face and voice. The early-appearing ability to match audio and visual cues based on phonetic information may involve innate detection of amodal invariants (Kuhl & Meltzoff, 1984). Our results suggest that, compared to phonetic matching, gender matching may require more experience and learning of face–voice relations that are arbitrary and culture-specific. It is not necessarily the case that young infants are better at phonetic matching than at gender matching. Rather, infants appear to focus on events that are salient for them given their stage of linguistic and social development.

REFERENCES

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., & Pickens, J. (1994). Amodal relations: The basis for intermodal perception and learning in infancy. In D. J. Lewkowicz & R. Lickliter (Eds.), *The development of intersensory perception: Comparative perspectives* (pp. 205–232). Hillsdale, NJ: Erlbaum.

Brown, E., & Perret, D. (1993). What gives a face its gender? *Perception*, **22,** 829–840.

Bruce, V., Burton, A., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R., & Linney, (1993). Sex discrimination: How do we tell the difference between male and female faces? *Perception*, **22,** 131–152.

Chronicle, E. P., Chan, M., Hawkings, C., Mason, K., Smethurst, D., Stallybrass, K., Westrope, K., & Wright, K. (1995). You can tell by the nose—Judging sex from an isolated facial feature. *Perception*, **24,** 969–973.

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

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

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

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

Gogate, L., Walker-Andrews, A. W., & Bahrick, L. (2001). The intersensory origins of word comprehension: An ecological–dynamic systems view. *Developmental Science*, **4,** 31–37.

Green, K., & Kuhl, P. K. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. *Perception & Psychophysics*, **50,** 534–536.

Hall, G., Lee, S., & Belanger, J. (2001). Young children's use of syntactic cues to learn proper names and count nouns. *Developmental Psychology*, **37,** 298–307.

Hirsch-Pasek, K., & Golinkoff, R. M. (1992). Skeletal supports for grammatical learning: What infants bring to the language learning task. In L. P. Lipcott & C. Rovee-Collier (Eds.), *Advances in infancy research* (Vol. 8., pp. 299–338). Norwood, NJ: Ablex.

Humphrey, K., & Tees, R. C. (1980). Auditory–visual coordination in infancy: Some limitations of the preference methodology. *Bulletin of the Psychonomic Society*, **16,** 213–216.

Jusczyk, P., & Bertoncini, J. (1988). Viewing the development of speech perception as an innately-guided learning process. *Language and Speech*, **31,** 217–238.

Kellman, P. J., & Arterberry, M. E. (1998). *The cradle of knowledge: Development of perception in infancy*. Cambridge, MA: MIT Press.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal development of speech in infancy. *Science*, **218,** 1138–1141.

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

Kuhl, P. K., Williams, K., Lacerda, B., Stevens, K. N., et al. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, **255,** 606–608.

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

Leinbach. M. D., & Fagot, B. (1993). Categorical habituation to male and female faces: Gender schematic processing in infancy. *Infant Behavior and Development*, **16,** 317–332.

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

Lewkowicz, D. (1998). Infants' response to the audible and visible properties of the human face: II. Discrimination of differences between singing and adult-directed speech. *Developmental Psychology*, **32,** 261–274.

Lewkowicz, D. (August 1999). Infants' perception of the audible, visible, and bimodal attributes of talking and singing faces. In *Proceedings of the Audio-Visual Speech Processing Conference, University of California, Santa Cruz.*

Liberman, A. M., & Mattingly, I. G. (1985). Motor theory of speech perception revised. *Cognition*, **21,** 1–36.

MacKain, K., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left hemisphere function. *Science*, **219,** 1347–1349.

McClure, E. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, **126,** 424– 453.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, **264,** 746–748.

Meltzoff, A. N., & Kuhl, P. K. (1994). Faces and speech: Intermodal processing of biologically relevant signals in infants and adults. In D. J. Lewkowicz & R. Lickliter (Eds.), *The development of intersensory perception: Comparative perspectives* (pp. 335–367). Hillsdale, NJ: Erlbaum.

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

Morrongiello, B. A. (1994). Effects of collocation on auditory–visual interactions and cross-modal perception in infants. In D. J. Lewkowicz & R. Lickliter (Eds.), *The development of intersensory perception: Comparative perspectives* (pp. 235–263). Hillsdale, NJ: Erlbaum.

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

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.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, **20,** 421–435.

Poulin-Dubois, D., Serbin, L., Kenyon, B., & Derbyshire, A. (1994). Infants' intermodal knowledge about gender. *Developmental Psychology*, **30,** 436–442.

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

Walker-Andrews, A. (1994). Taxonomy for intermodal relations. In D. J. Lewkowicz & R. Lickliter (Eds.), *The development of intersensory perception: Comparative perspectives* (pp. 39–55). Hillsdale, NJ: Erlbaum.

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

Walton, G. E., & Bower, T. G. R. (1993). Amodal representation of speech in infants. *Infant Behavior and Development*, **16,** 233–243.

Werker, J. F., & Tees, R. C. (1992). The organization and reorganization of human speech perception. *Annual Review of Neuroscience*, **15,** 377–402.

Received February 21, 2001; revised June 25, 2001