ijbd

International Journal of Behavioral Development
37(2) 84–89

DOI: 10.1177/0165025412467583
ijbd.sagepub.com
SAGE

Six-month-old infants match other-race faces with a non-native language

Lesley Uttley,[1] Anne Hillairet de Boisferon,[2] Eve Dupierrix,[2] Kang Lee,[3] Paul C. Quinn,[4] Alan M. Slater,[5] and Olivier Pascalis[1,2]

Abstract

Early in life, infants possess an effective face-processing system which becomes specialized according to the faces present in the environment. Infants are also exposed to the voices and sounds of caregivers. Previous studies have found that face–voice associations become progressively more tuned to the types of association most prevalent in the environment. The present study investigated whether 6-month-old infants associate own-race faces with their native language and faces from a different race with a non-native language. Infants were presented with pictures of own- and other-race faces simultaneously, with a native or non-native language in a habituation paradigm. Results indicate that 6-month-olds are able to match other-race faces to a non-native language.

Keywords

face, infant cognition, perception

There is a general consensus that we become face experts (Carey, 1992) and that the immature face processing system that is present at birth develops with experience. Newborns prefer face-like stimuli to non-face stimuli and categorize/recognize familiar and unfamiliar faces (see Lee, Anzures, Quinn, Pascalis, & Slater, 2011, for a review). The face processing system will then develop rapidly, with infants exposed to more and more faces in their environment. By 3 months of age, infants show a visual preference for faces that match the gender of their primary caregiver (Quinn, Yahr, Kuhn, Slater, & Pascalis, 2002) and for faces of their own-ethnic group (Bar-Haim, Ziv, Lamy, & Hodes, 2006; Kelly et al., 2005; Kelly, Liu et al., 2007), demonstrating the importance of experience with faces from the infant's visual environment and how they impact upon the development of the face processing system.

Infants are exposed to faces at the same time that they are exposed to the voices of their caregivers and therefore need to integrate information from the two modalities. This multisensory perception of faces and voices allows for creating categories of gender and language, but also learning identity: the arbitrary relationship of the sound of someone's voice and their face. Sai (2005) demonstrated that newborns only recognize their mother's face if a postnatal exposure to the mother's voice–face combination was available. The mother's face is then learned in conjunction with the mother's voice, which has been heard during gestation (Kisilevsky et al., 2003); if the infant is denied the auditory input of the mother's voice after birth, recognition of the mother's face is not demonstrated. Further evidence of early inter-sensory perceptual abilities has been found in young infants. For example, infants are sensitive to bimodal events in terms of temporal synchrony from birth (Slater, Quinn, Brown, & Hayes, 1999), and are able to match intermodal phonetic information from 2 months of age (Patterson & Werker, 2003). Nevertheless, a delay is reported in matching of more arbitrary face–voice relations that allow the infant to form multimodal categories such as species or gender. Some studies seem to indicate the presence of sensitivity to a face–voice relationship for gender between 4 and 6 months of age, and a more robust perception only among 9-month-old infants (Patterson & Werker, 2002; Walker-Andrews, Bahrick, Raglioni, & Diaz, 1991). Vouloumanos, Druhen, Hauser, and Huizink (2009) found that infants at the age of 5 months matched the faces and voices of humans and animals, potentially indicating an intermodal understanding of species-specific information at this age.

During the first year of life, both the face and language system develop from broad non-specific systems to more specialized native-language and human-tuned systems (Nelson, 2001; Werker & Tees, 1984). The ontogeny of phonemic perception is characterized by a decline in the discrimination of speech sounds not present in one's native language. This specialization occurs between 6 and 12 months of age and is critically dependent on perceptual exposure to native, relative to non-native, phonemic contrasts (Werker & Tees, 2005). Before 9 months of age, infants can discriminate faces from different species (Pascalis, de Haan, & Nelson, 2002; Simpson, Varga, Frick, & Fragaszy, 2011) or race (Kelly, Quinn et al., 2007), but at around 9 months of age their recognition abilities are limited to own species and own race. Such perceptual narrowing has also been observed in multisensory processes. For example, Lewkowicz and Ghazanfar (2006) showed that 4- and 6-month-olds, but not 8- and 10-month-olds, exhibited intersensory matching of monkey faces and monkey vocalizations. The system not only narrows toward our own-species, but also toward our native

[1] University of Sheffield, UK
[2] Université Pierre Mendès, France
[3] University of Toronto, Canada
[4] University of Delaware, USA
[5] University of Exeter, UK

Corresponding author:
Olivier Pascalis, Laboratoire de Psychologie et NeuroCognition, Université Pierre Mendès, BP 47 38040, Grenoble, Cedex 9, France.
Email: Olivier.Pascalis@upmf-grenoble.fr

language. Pons, Lewkowicz, Soto-Faraco, and Sebastián-Gallés (2009) found that the intersensory response to a non-native phonetic contrast narrows between 6 and 11 months of age, suggesting that the perceptual system becomes increasingly more tuned to key native language audio-visual correspondences.

Altogether, these research studies show the existence of perceptual intersensory abilities on arbitrary face–voice relations at around 6 months of age, and a modulation by the experience acquired during the first year. The perceptual abilities of face–voice relations might still develop for familiar categories for which infants have daily experience (e.g., for their own species), while declining for categories for which infants have little experience with (e.g., other species).

If 6-month-old infants are able to form multimodal categories, like gender and species based on face–voice associations, are they able to form other multimodal categories such as race? In other words, do infants expect a face of a particular race to speak a specific language? The present study investigated whether 6-month-old infants associate an own-race face with their native language and a face from a different race with a non-native language? We hypothesized that Caucasian, English-speaking, infants raised in monolingual families would look longer to matching conditions in which Caucasian faces were accompanied by English language and Chinese faces were accompanied by Mandarin language, compared to mismatching conditions in which Caucasian faces were accompanied by Mandarin language and Chinese faces were accompanied by English language. Although Chinese faces and Mandarin language feature as the other-race stimuli in contrast to our Caucasian faces and English language soundtrack, the implication is not that infants are capable of matching Chinese faces to the Mandarin language as a matter of intuition. It is possible that any other-race faces would be ''matched'' with any other language that is non-native to the infant and therefore deemed as unfamiliar.

Experiment 1

Participants

Thirty-six 6-month-old Caucasian infants raised in monolingual families were included in the final sample (19 males, mean age = 185 days, $SD = 9.6$). A further 11 infants were eliminated from the analysis due to fussiness ($n = 9$), parental interference ($n = 1$), or experimenter error ($n = 1$).

Stimuli

The visual stimuli consisted of static, colour images of individual Caucasian (4) and Chinese female (4) adult faces (see Figure 1 for examples). The images were approximately 10 cm in height when presented on a computer monitor and were rated by adults as being similar in attractiveness and expression.

The sound stimuli consisted of two female voice recordings in English or Mandarin language. A native English-speaking female was recorded saying ''Hello baby, look at me. Peek-a-boo! Look at me.'' and a native Mandarin-speaking female saying ''Xiao bao-bao, kankanwoh, Meh! Kankanwoh.'' (Mandarin sound translation of the English). In order to ensure that the English recording was recognized by the participants as their native language, we recruited a local English-speaking female. Both speakers were asked to speak clearly and in a friendly tone as though speaking to a baby.

Figure 1. Examples of stimuli used in all experiments, presented in conjunction with a matching or mismatching language soundtrack.

Procedure

An infant control procedure was used whereby one face was presented for a maximum trial duration of 60 seconds or until the infant averted its gaze from the stimuli for 2 seconds or more. There were four trials: two were with Chinese female faces and two were with Caucasian female faces. The visual stimuli were presented simultaneously with the sound stimuli. One group of infants was assigned to the matching condition: They saw two Caucasian faces with English soundtracks and two Chinese faces with Mandarin soundtracks. A second group was assigned to the mismatching condition: They saw two Caucasian faces with Mandarin soundtracks and two Chinese faces with English soundtracks. The order of test conditions was counterbalanced between subjects.

Analysis of data

Looking-time data were analysed using a two-way mixed ANOVA. Face–voice matching was a between-subjects factor (participants were either in a condition with all matching face–voice stimuli or all mismatching face–voice stimuli) and the race of the face (Caucasian or Chinese) was a within-subjects factor.

Results

An ANOVA showed a significant effect of face–voice matching, $F(1, 34) = 20.6$, $p < .001$. Infants in the face–voice matching condition looked significantly longer at the stimuli ($M = 60.16$ seconds) than infants in the mismatching condition ($M = 30.74$ seconds) (see Figure 2). No main effect for race of the face, $F(1, 34) = 1.36$, $p = .25$, nor race of the face × face–voice matching interaction was found, $F(1, 34) = 0.72$, $p = .79$.

Discussion

The results show that 6-month-olds who viewed Caucasian faces with an English soundtrack and Chinese faces with a Mandarin soundtrack looked longer at the faces than infants who were in the mismatching condition. The findings suggest that infants have developed an intermodal concept of race and language for both their native and a non-native race at 6 months of age, preferring to look at combinations of face and voice samples which go together. This is in line with other reports in the developmental literature looking at infants' abilities to form intermodal concepts across categories such as gender (Walker-Andrews et al., 1991) and age (Bahrick, Netto, & Hernandez-Reif, 1998). However, Experiment 1 was a between-

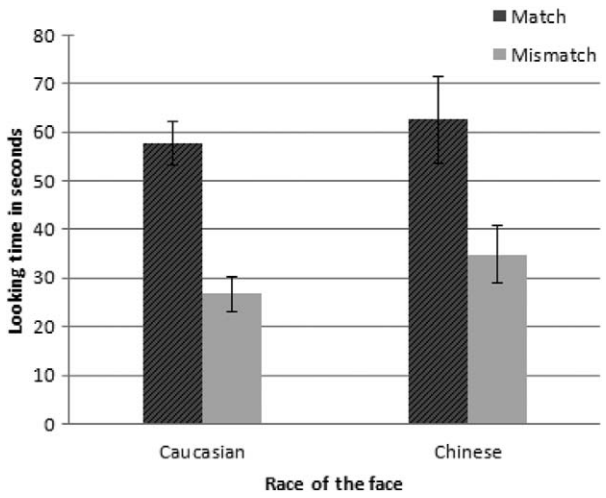

Figure 2. Mean looking-times in seconds (±SEM) at Caucasian and Chinese faces with matching or mismatching voices. One group of 6-month-olds was assigned to the match condition, and a different group of infants was assigned to the mismatch condition.

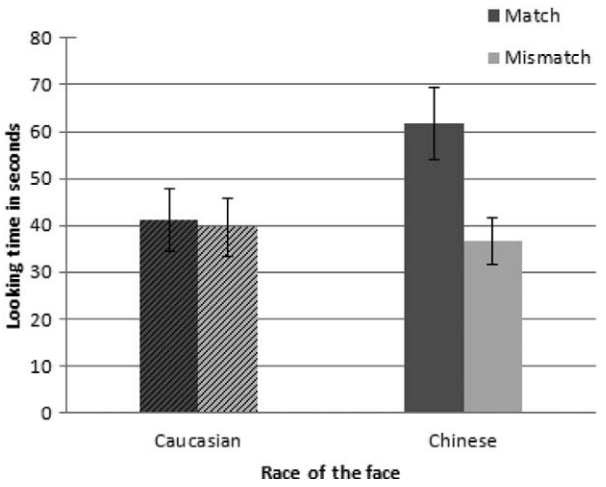

Figure 3. Mean looking-time in seconds (±SEM) at Chinese and Caucasian faces with matching or mismatching voices. One group of infants was tested with Caucasian faces, and a different group was tested with Chinese faces.

participants design, and therefore Experiment 2 was performed to determine if the same behaviour would be observed with a within-subjects design.

Experiment 2

Participants

Fifty-six 6-month-old Caucasian infants raised in monolingual families (31 males, mean age = 187 days, $SD = 4.5$) participated in the second experiment. A further 12 infants were eliminated from the analysis due to fussiness ($n = 10$), equipment failure ($n = 1$), or experimenter error ($n = 1$).

Stimuli

The stimuli were identical to Experiment 1.

Procedure

An infant control procedure was used as previously. Infants were allocated to either a "Caucasian" or a "Chinese" condition; however, all infants received exposure to both matching and mismatching face–voice pairs. In the "Caucasian" condition, infants were presented with four Caucasian female faces presented one at a time and an accompanying soundtrack that was in either English language (for two trials, matching condition) or Mandarin language (for the two remaining trials, mismatching condition). Similarly, in the "Chinese" condition, infants were presented with four different Chinese female faces which were paired with English (mismatching condition) or Mandarin language (matching condition). The order of presentation of matching and mismatching conditions was pseudo-randomized.

Analysis of data

Looking time data were analysed using a two-way mixed ANOVA. The race of the face (Caucasian vs. Chinese) was a between-subjects factor and the face–voice matching (matching vs. mismatching) was a within-subject factor.

Results

The ANOVA showed no significant effect of the race of the face, $F(1, 54) = 1.14$; $p = .29$. Infants looked the same amount of time at both kinds of faces ($M_{\text{Chinese}} = 49.13$ seconds, $M_{\text{Caucasian}} = 40.36$ seconds). There was a significant effect of matching, $F(1, 54) = 9.96$, $p < .01$, indicating that infants looked longer at the matching face than at the mismatching face (see Figure 3), but this effect is mainly due to the matching effect in the Chinese-face condition, as shown by the significant interaction between race of the face and face–voice matching, $F(1, 54) = 7.89$, $p < .01$. A Scheffé test revealed that infants looked significantly longer at the Chinese face when hearing Mandarin than when listening to English ($p < .01$). No difference was observed for the Caucasian faces ($p = 1$).

Comparison of Experiments 1 and 2

In order to determine if the change in the design affected the results we compared with a *t*-test the mean looking time for each condition (see Table 1). The same pattern of looking time is observed in both experiments for both Caucasian and Chinese faces in the face–voice matching and mismatching conditions.

Discussion

These results show that 6-month-olds looked longer at Chinese faces when presented with Mandarin language than when associated with English language. The findings suggest again that infants possess an intermodal concept of race and language. Nevertheless, the outcome that infants did not look longer at Caucasian faces when either English or Mandarin language was heard could suggest that they do not have an intermodal concept of race and language for their own race. Another explanation could be that to young infants, new faces speaking a novel language capture and maintain more of their attention. Overall, when this

Table 1. Mean looking-time in seconds (standard deviations in brackets) for Caucasian and Chinese faces in Experiments 1 and 2 for the face–voice matching and mismatching conditions.

	Face–voice matching		Face–voice mismatching	
	Caucasian face	Chinese face	Caucasian face	Chinese face
Experiment 1	57.64 (18.80)	62.68 (37.72)	26.70 (15.82)	34.76 (24.99)
Experiment 2	41.09 (37.21)	61.59 (39.46)	39.63 (33.92)	36.67 (24.92)
t-test	$p = .16$	$p = .92$	$p = .13$	$p = .83$

methodology is used, infants find it more plausible that an unfamiliar face type belongs with an unfamiliar language. In contrast, when one element is familiar (either an own-race face or native language), then interest is equal.

Experiment 3

There is the possibility that the infants are not matching other-race faces with other languages, but that they are identifying any sound which is different from their native language with other-race faces. This could mean that infants could be matching non-speech sounds with other-race faces, and are therefore less likely to be tapping into their sensitivity to race and language per se. In order to identify that the effect of intermodal matching for other-race faces is not merely an artefact of matching any non-native sounds with other-race faces, a control study was required to establish that the infants will only look longer at other-race faces when paired with a non-native language, and not when paired with any non-native sounds.

Participants

A total of 18 6-month-old Caucasian infants raised in monolingual families participated in the study. Of the total number of infants who participated in the study, two were excluded from analysis due to fussiness. The final sample therefore consisted of 16 6-month-old infants: 9 males, 7 females (mean age = 189 days, $SD = 4.2$).

Stimuli

The face stimuli were identical to those used in the previous experiments. However, the sound stimuli consisted of backwards speech made from the soundtracks used in the previous experiments (English and Mandarin). Therefore, the duration, pitch, and volume of the sound stimuli were the same as those in the previous experiments; however, they no longer contained any properties of natural speech or language.

Procedure

Using the same procedure as in Experiment 1, a pilot study showed that infants did not pay enough attention to faces while backward speech was playing to allow us to collect valid data. We therefore decided to use a preferential looking-time procedure. There were four trials, during which each infant was shown a different pair of Caucasian and Chinese female faces. Images were displayed side-by-side, separated by an 11 cm gap. The different pair combinations were randomized across participants. The left/right positioning of Caucasian and Chinese faces was counterbalanced across infants on the first trial and reversed on the following trial. The visual stimuli were presented simultaneously with the sound stimuli. Each infant was presented with backward English language (two trials) and backward Mandarin language soundtracks (two trials). The presentation order of the English and Mandarin soundtracks was also counterbalanced across participants. Infants' gaze was centred at the beginning of each trial. The experimenter began recording looking-time when the infant looked at one of the two stimuli. The trial ended after a cumulative 10 seconds of looking-time duration. Looking time for both kinds of faces was summed across the two speech conditions.

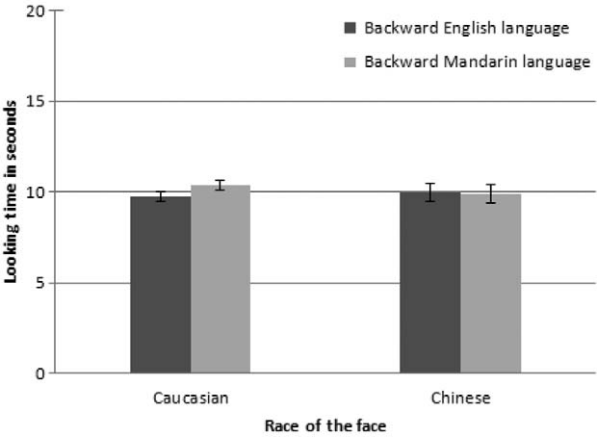

Figure 4. Mean looking-time in seconds (±SEM) to Caucasian and Chinese faces with backward English and Mandarin languages.

Analysis of data

Looking-time data were analysed using a repeated measure ANOVA with the race of the face (Caucasian vs. Chinese) and sound stimuli (backward English vs. backward Mandarin) as within-subject factors.

Results

An ANOVA found no significant difference in looking time between Caucasian and Chinese faces when hearing backwards English, $F(1, 15) = 0.19$, $p > 0.05$ ($M_{\text{Caucasian}} = 9.75$ seconds; $M_{\text{Chinese}} = 10.00$ seconds) or when hearing backwards Mandarin, $F(1, 15) = 1.03$, $p > 0.05$ ($M_{\text{Caucasian}} = 10.36$ seconds; $M_{\text{Chinese}} = 9.89$ seconds) (see Figure 4).

Discussion

The finding that infants did not look longer at either Caucasian or Chinese faces following backwards speech shows that infants do not associate any unusual sound with a new category of faces. This result reinforces the data from the previous experiments that infants do not merely match unfamiliar sounds with other-race faces, but match other-race faces with other *languages*. The backwards speech soundtracks contained the same volume, pitch, and fundamental frequency as the speech soundtracks but they lacked the audible properties of natural language. This might explain why,

in this last experiment, infants looked at the stimuli for only 10 seconds each on average.

General discussion

First, the results of the three experiments consistently show 6-month-old infants to be able to match other-race faces to a non-native language. There are two possible explanations: First, 6-month-old infants may have a multimodal representation of individuals from their native group and present a preference for a novel multisensory stimulus: a novel face *belonging* to a novel language. The second interpretation is that performance could reflect a simple novelty effect. The amount of novelty represented when both face and voice are new could explain the enhanced infants' attention to Chinese faces. In contrast, when one of the elements is familiar, an own-race face or English language, infants' attention may be less well maintained.

However, with regard to the results of Experiment 3, it seems that the increase of looking time to the Chinese face–voice cannot only be triggered by the "novelty" of stimuli. The ability of 6-month-olds to match other-race faces to non-native language was not simply due to the sound itself being different from language as infants know it. Indeed, infants did not match other-race faces with backwards speech, and exhibited a null preference for both Caucasian and Chinese faces. This result suggests that the 6-month-old infants tested in this set of experiments were matching race to *language* rather than to unfamiliar sounds per se. The finding provides evidence that infants at this age have categorical perception for language versus nonsensical human sound, and have expectations about human faces being associated with language, whether it be native or non-native. This is consistent with Vouloumanos, Druhen, Hauser, & Huizinik's (2009) findings that infants were able to match the face and vocalizations of humans and monkeys according to species. Vouloumanos et al. concluded that the matching ability does not appear to be based on simple associative learning or prior experience, and also that infants actually possess expectations about the sources of vocalizations (Bonatti, Frot, Zangl, & Mehler, 2002).

Our findings for infants' abilities to match own-race faces to native language were less consistent. We observed them in Experiment 1, but not in Experiment 2. We initially concluded that infants have developed an intermodal concept of race and language for their native language. It could be that the intermodal representation is not stable at 6 months of age and is sensitive to the familiarity level of the stimuli and to the task used.

Our results are generally consistent with the existing literature inasmuch as the evidence of face–voice matching seems to emerge at around 6 months of age for gender and becomes more robust around 9 months of age (Patterson & Werker, 2002; Walker-Andrews et al., 1991). Nevertheless, it is surprising that young infants at least discriminate new face and new language in Experiments 1 and 2, categorize language versus nonsensical human sound in Experiment 3, but do not match own-race faces to native language despite an overwhelming experience with faces of their own race and with their native language. When dynamic stimuli are used it was demonstrated that infants pay a lot of attention to mouth movement/sound synchrony (Lewkowicz & Hansen-Tift, 2012; Patterson & Werker 2003). The presentation of Caucasian faces associated with English language could generate greater expectations, as they are familiar categories, for synchrony between produced sounds and lips' motion. It is possible that a static picture when speech is presented is not the best set-up and creates interferences for own-race faces and native language more than for other-race faces and non-native language. The only way to solve this issue would be to conduct further investigations with sound-synchronized videos.

Funding

This study was supported by a grant from National Institutes of Health Grant R01 HD046526.

References

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

Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R. M. (2006). Nature and nurture in own-race face processing. *Psychological Science*, *17*, 159–163.

Bonatti, L., Frot, E., Zangl, R., & Mehler, J. (2002). The human first hypothesis: Identification of conspecifics and individuation of objects in the young infant. *Cognitive Psychology*, *44*, 388–426.

Carey, S. (1992). Becoming a face expert. *Philosophical Transactions of the Royal Society of London*, *335*, 95–103.

Kelly, D. J., Liu, S., Ge, L., Quinn, P. C., Slater, A. M., Lee, K., & Pascalis, O. (2007). Cross-race preferences for same-race faces extend beyond the African versus Caucasian contrast. *Infancy*, *11*, 87–95.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science*, *18*, 1084–1089.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., & Pascalis, O. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science*, *8*, F31–F36.

Kisilevsky, B. S., Hains, S. M. J., Lee, K., Xie, X., Huang, H., Ye, H. H., & Wang, Z. (2003). Effects of experience on fetal voice recognition. *Psychological Science*, *14*, 220–224.

Lee, K., Anzures, G., Quinn, P. C., Pascalis, O., & Slater, A. (2011). Development of face processing expertise. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *Oxford handbook of face processing* (pp. 753–778). Oxford, UK: Oxford University Press.

Lewkowicz, D. J., & Ghazanfar, A. A. (2006). The decline of cross-species intersensory perception in human infants. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 6771–6774.

Lewkowicz, D. J., & Hansen-Tift, A. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 1431–1436.

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

Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science*, *296*, 1321–1323.

Patterson, M. L., & Werker, J. F. (2002). Infants' ability to match dynamic phonetic and gender information in the face and voice. *Journal of Experimental Child Psychology*, *81*, 93–115.

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, *6*, 191–196.

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy.

Proceedings of the National Academy of Sciences of the United States of America, *106*, 10598–10602.

Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *Perception*, *31*, 1109–1121.

Sai, F. Z. (2005). The role of the mother's voice in developing mother's face preference: Evidence for intermodal perception at birth. *Infant and Child Development*, *14*, 29–50.

Simpson, E., Varga, K., Frick, J., & Fragaszy, J. (2011). Infants experience perceptual narrowing for nonprimate faces. *Infancy*, *16*, 318–328.

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

Vouloumanos, A., Druhen, M. J., Hauser, M. D., & Huizink, A. T. (2009). Five-month-old infants' identification of the sources of vocalizations. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 18867–18872.

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

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49–63.

Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology*, *46*, 233–251.