

Infancy, 20(6), 661–674, 2015

Copyright © International Congress of Infant Studies (ICIS)

ISSN: 1525-0008 print / 1532-7078 online

DOI: 10.1111/infa.12088

WILEY Blackwell

Perception of Multisensory Gender Coherence in 6- and 9-Month-Old Infants

Anne Hillairet de Boisferon and Eve Dupierrix

*Laboratoire de Psychologie et NeuroCognition
CNRS - UMR5105, Université Grenoble Alpes*

Paul C. Quinn

*Department of Psychological and Brain Sciences
University of Delaware*

Hélène Løevenbruck

*Laboratoire de Psychologie et NeuroCognition
CNRS - UMR5105, Université Grenoble Alpes*

David J. Lewkowicz

*Department of Communication Sciences & Disorders
Northeastern University*

Kang Lee

*Institute of Child Study
University of Toronto*

Olivier Pascalis

*Laboratoire de Psychologie et NeuroCognition
CNRS - UMR5105, Université Grenoble Alpes*

Correspondence should be sent to Anne Hillairet de Boisferon, Department of Communication Sciences & Disorders, Northeastern University, 360 Huntington Ave., 226FR, Boston, MA 02115. E-mail: anneh2b@gmail.com

[Corrections added on 8 July 2015, after first online publication: The copyright line in this article was changed to reflect the society's current name.]

One of the most salient social categories conveyed by human faces and voices is gender. We investigated the developmental emergence of the ability to perceive the coherence of auditory and visual attributes of gender in 6- and 9-month-old infants. Infants viewed two side-by-side video clips of a man and a woman singing a nursery rhyme and heard a synchronous male or female sound track. Results showed that 6-month-old infants did not match the audible and visible attributes of gender, and 9-month-old infants matched only female faces and voices. These findings indicate that the ability to perceive the multisensory coherence of gender emerges relatively late in infancy and that it reflects the greater experience that most infants have with female faces and voices.

One of the most salient and socially relevant multisensory categories specified by faces and voices is gender. Even though gender can be specified by separate facial and vocal attributes (e.g., Brown & Perrett, 1993; Coleman, 1971), adults typically expect that male faces go with male voices and that female faces go with female voices, and perceive the visual and auditory attributes of an individual's gender as part of a unitary and coherent entity (Barenholtz, Lewkowicz, Davidson, & Mavica, 2014; Kamachi, Hill, Lander, & Vatikiotis-Bateson, 2003).

Research has determined that the ability to perceive various forms of multisensory coherence in faces and voices begins to emerge during infancy. For example, newborn infants can recognize their mother's face following postnatal exposure to her voice and face (Sai, 2005), and newborn and 3-month-old infants can associate an unfamiliar person's face and voice after familiarization with both (Brookes et al., 2001; Coulon, Guellai, & Streri, 2011; Guellai, Coulon, & Streri, 2011). Similarly, studies show that infants can perceive the multisensory coherence of the audible and visible attributes of speech syllables (e.g., Kuhl & Meltzoff, 1982; Patterson & Werker, 1999) by 4 months of age, and the equivalence of audible and visible fluent speech by the end of the first year of life (Lewkowicz, Minar, Tift, & Brandon, 2015).

Given that multisensory perception of various aspects of faces and voices emerges in infancy, and that gender is one of the most frequently experienced perceptual categories in early development, perception of the multisensory coherence of gender may also emerge in infancy. This possibility is all the more likely given that infants can detect the gender of voices by 2 months of age (Jusczyk, Pisoni, & Mullennix, 1992; Miller, 1983; Miller, Younger, & Morse, 1982) and begin to categorize human faces based on gender between 3 and 12 months of age, although with certain processing advantages for female over male faces (Cornell, 1974;

Fagan & Singer, 1979; Leinbach & Fagot, 1993; Quinn, Yahr, Kuhn, Slater, & Pascalis, 2002; Ramsey, Langlois, & Marti, 2005).

Several previous studies have reported that infants can perceive the multisensory coherence of gender (see Ramsey et al., 2005, pp. 219–220; for a review). For example, Walker-Andrews, Bahrick, Raglioni, and Diaz (1991) found that 6-month-old infants matched faces and voices based on gender, although this was only the case in half the trials. In addition, to ensure that the faces and voices were synchronized, Walker-Andrews et al. dubbed each speaker's voice onto the same speaker's video. This procedure allowed for the possibility that matching was partially due to infant detection of the idiosyncratic relation between the specific faces and voices, an observation which is supported by findings that adults (Kamachi et al., 2003; Rosenblum et al., 2002) as well as 6-month-old infants (Trehub, Plantinga, & Brcic, 2009) can match dynamic auditory and visual speaker identity cues.

Subsequent studies only found evidence of the ability to perceive the multisensory coherence of gender starting either at 8 months of age (Patterson & Werker, 2002) or as late as 9–12 months of age (Poulin-Dubois, Serbin, Kenyon, & Derbyshire, 1994). One especially interesting finding reported in several studies is that between 8 and 12 months of age, infants only exhibit perception of the multisensory coherence of female faces and voices (Noles, Kayl, & Rennels, 2012; Poulin-Dubois et al., 1994). Another noteworthy finding is that it is not until 18 months of age that infants can recognize the multisensory coherence of gender for males as well as females (Poulin-Dubois, Serbin, & Derbyshire, 1998).

The seeming discrepancy in the specific age when the ability to perceive the multisensory coherence of gender first emerges is probably due to the fact that the particular nature of the presented faces and vocalizations has varied across different studies. Specifically, some studies have presented animated faces (Noles et al., 2012; Patterson & Werker, 2002; Walker-Andrews et al., 1991), whereas others have presented pictures of static faces (Poulin-Dubois et al., 1998, 1994). It is possible that the earlier emergence of the ability to perceive the multisensory coherence of gender with animated faces is due to the fact that infants can more easily perceive the physiognomic aspects of such faces (Xiao et al., 2014). Similarly, some studies have presented recordings of fluent speech (Noles et al., 2012; Poulin-Dubois et al., 1998, 1994; Walker-Andrews et al., 1991), whereas other studies have presented isolated vowels (Patterson & Werker, 2002). The former may make the detection of gender *per se* easier because continuous utterances not only provide fundamental frequency and formant cues (as do isolated syllables) which usually

distinguish between male and female voices (Gelfer & Mikos, 2005; Ladefoged & Broadbent, 1957; Peterson & Barney, 1952; Sachs, Lieberman, & Erickson, 1973), but they also provide cues related to intonation, stress, duration, and vocal breathiness. Any of these cues could potentially signal the gender of a person (see, Klatt & Klatt, 1990; Van Borsel, Janssens, & De Bodt, 2009; for examples on vocal breathiness).

The various issues raised by the specific types of stimulus materials used in past research justify further investigation. Based on previous findings, we predicted that the ability to perceive multisensory gender coherence is most likely to emerge between 6 and 9 months of age (Patterson & Werker, 2002; Poulin-Dubois et al., 1994; Walker-Andrews et al., 1991). Therefore, we investigated the ability of 6- and 9-month-old infants to match auditory and visual gender information inherent in dynamic displays of male and female faces that could be seen and heard singing. We chose singing as opposed to continuous speech for three reasons. First, songs are more prosodically varied and infants are known to be highly sensitive to prosody both in the auditory (Fernald, 1985) and visual domains (Kim & Johnson, 2014; Nakata & Trehub, 2004). Second, the sensitivity of infants to song prosody is likely to facilitate their detection of the multisensory coherence of gender. Finally, singing, unlike fluent speech, spontaneously elicits similar rhythms in different singers, which is not the case when speakers are asked to utter a monologue.

The experiment consisted of two trials where infants saw side-by-side singing faces, one a male face and the other a female face and, at the same time, heard a male or a female person singing. If the infants could detect the gender-based correspondence, we expected them to look longer at the face whose gender matched the gender of the accompanying voice than at the face that did not match it. An additional expectation was that infants might only exhibit successful perception of the multisensory coherence of female faces and voices. This prediction is based on studies that have found that 3- to 4-month-old infants reared by a female primary caregiver exhibit a spontaneous preference for female over male faces (Quinn et al., 2002) and that infants are more efficient at categorizing female as opposed to male faces (Ramsey et al., 2005; Quinn et al., 2002; Younger & Fearling, 1999). As indicated earlier, prior studies have also shown that, starting at 8–9 months of age, infants exhibit evidence of intersensory matching of gender for female faces and voices (Patterson & Werker, 2002; Poulin-Dubois et al., 1994). Crucial from the standpoint of the current study, however, is the fact that studies involving infants older than 6 months only presented single auditory and visual isolated syllables or static faces. Because of this methodological limitation, it is not known

whether the results from these studies generalize to infant perception of the multisensory coherence of dynamic gender cues. Answering this question is important because dynamic gender cues are more representative of an infant's typical ecological setting than isolated syllables or static faces.

METHOD

Participants

Forty-eight 6-month-old infants ($M_{\text{age}} = 193$ days, $SD = 7$ days, 182–215 days; 26 females) and 47 9-month-old infants ($M_{\text{age}} = 285$ days, $SD = 7$ days, 271–299 days; 21 females) were included in the final analysis. All participants were healthy full-term infants. They were recruited from the maternity ward of the Centre Hospitalier Universitaire, Grenoble, France. Eighteen additional infants were tested but eliminated from the analysis due to technical problems (6-month-olds, $N = 1$; 9-month-olds, $N = 2$), crying (6-month-olds, $N = 1$), or side bias (defined as greater than 95% of looking time to one side of the display; 6-month-olds, $N = 10$; 9-month-olds, $N = 4$).

Stimuli

First, we video-recorded the face and voice of a female adult singing a traditional French nursery rhyme in French and used this person's recording as a template for the way other actors should sing it. After playing this template video to each of six adult actors (three females and three males) for several minutes, we video-recorded the actors while they sang the nursery rhyme in a manner as similar to the model singer as possible. During the video recording, the actors looked directly at the video camera, maintained a neutral facial expression, and sang at the same tempo. We edited the videos to make sure that they started with the same opening of the mouth. To draw the attention of the infants to the physiognomic features of the faces, only faces were presented and culture-specific visual information (i.e., makeup, jewelry) was removed. Figure 1 shows an example of the faces presented in the experiment.

The final stimulus set consisted of six different faces (20 cm in height and 15 cm in width) and six different accompanying voices. Three of the faces and voices were male and three were female. Twelve different pairs of male and female face–voice combinations were created (six male–female face combinations associated with a male or a female voice). The voice presented during the presentation of each male–female face pair was not the voice of either of the actors seen in the video. Importantly, we dubbed

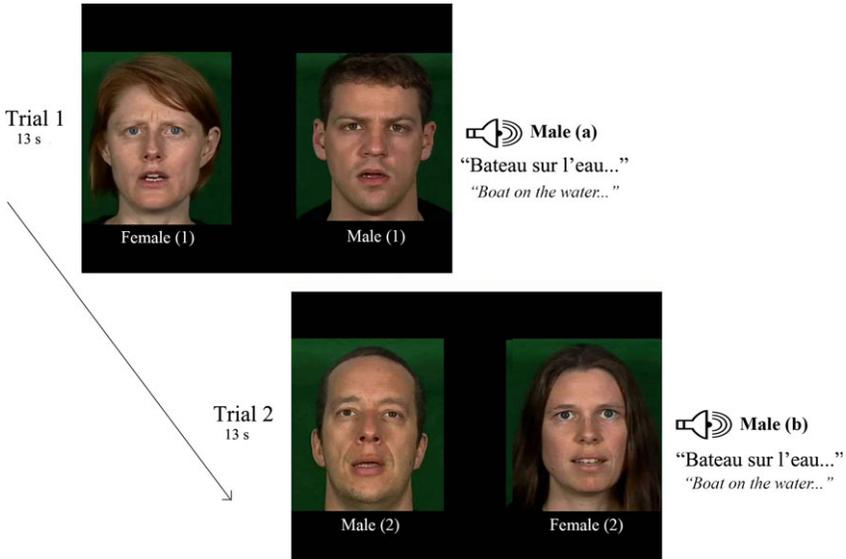


Figure 1 Schematic representation of the procedure. Infants saw a male and female face singing at the same time and heard the accompanying song during each of the two paired-preference trials. The male voice condition is shown.

another person's singing onto the videos to ensure that infants could not make the match based on speaker identity, or idiosyncrasies in visible and audible articulatory or respiratory patterns.

A perceptual test was carried out to measure the perceived degree of synchrony of each of the 24 individual videos (each actor–voice combination). Twenty French participants were tested (10 males, $M_{\text{age}} = 34.3$, $SD = 9.2$). The instruction was to watch each video carefully and evaluate the degree of synchrony between the image and the sound, on a Likert scale from 1 (very poor synchrony, long lag between image and sound) to 7 (perfect synchrony, no observable lag). Results showed that the degree of agreement between observers was high, as revealed by the intraclass correlation coefficient of 0.88 (95% confidence interval 0.78–0.94). The fact that none of the facial movements was perfectly synchronized with the voices is inherent to this procedure, but the fact that none of the faces of a pair was better synchronized than its counterpart with the sound ensures that the potential matching cannot be based on synchronization cues ($M_{\text{males}} = 5.1$, $SD = 0.8$; $M_{\text{females}} = 4.9$, $SD = 1.4$; $t(11) = 0.7$, $p = .50$, paired two-samples t -test).

Procedure

Infants were tested individually in a quiet room. They were seated on their caregiver's lap, 60 cm away from a 22-inch monitor that showed the videos. E-Prime software (Psychology Software Tools, Pittsburgh, PA, USA) was used to present the videos. The experiment consisted of two 13-sec trials during which infants saw the side-by-side video clips of a man and a woman singing a nursery rhyme and heard either a male or female sound track (Figure 1). The two faces were separated by a 15-cm gap. For each infant, the gender of the voice was the same in both trials. The male or female voices were randomly assigned to infants and differed in terms of their identity on each trial. Also, different pairs of male and female faces were presented across the two test trials. We deliberately varied the specific identity of the voices and faces across the two test trials to maximize our ability to make generalizations regarding infant ability to perceive the multisensory coherence of gender. The side of gender presentation (male or female) was counterbalanced across infants on the first trial and reversed on the following trial.

All parents were instructed to fixate centrally above the screen and to remain quiet during testing. A low-light video camera, located above the stimulus-presentation monitor, was used to record infant looking behavior. This video recording was subsequently digitized and analyzed frame by frame by two independent observers (interobserver agreement calculated as a Pearson r was 0.94). The dependent measure was the total duration of looking directed to each face.

RESULTS

Infants spent an average of 95.2% of their time looking at the faces. To determine whether infants perceived multisensory gender coherence, we calculated the mean proportion of total looking time (PTLT) that each infant directed at the matching face over the two test trials. For this calculation, we divided the total amount of looking at the matching face by the total amount of looking at both faces during each trial and then averaged those two proportions over the two test trials. If infants perceived multisensory gender coherence, they should exhibit greater looking at the face that matched the heard voice (i.e., PTLT to the matching face superior to the 50% chance level).

An analysis of variance (ANOVA) was first conducted on the PTLT directed at the matching face to determine whether voice gender (male, female), infant gender (male, female), or age (6-month-olds, 9-month-olds)

affected looking behavior. Results indicated that only the gender of the voice affected responsiveness, $F(1, 87) = 5.14$, $p < .05$, $\eta_p^2 = 0.06$. As expected, infants looked longer at the matching face in presence of a female voice ($M = 53.8\%$, $SD = 8.5\%$) than a male voice ($M = 49.4\%$, $SD = 10.3\%$). Neither age $F(1, 87) = 0.09$, $p = .77$, $\eta_p^2 = 0.001$, nor infant gender, $F(1, 87) = 0.02$, $p = .89$, $\eta_p^2 < 0.001$, nor any of the interactions between factors reached significance.

We further investigated infant ability to perform audiovisual matching by comparing mean PTLTs directed at the matching face against chance. One-sample t -tests against chance were performed at each age, for male and female voices, in order to compare more directly our results to previous studies. The data for these comparisons can be seen in Figure 2. The analyses revealed differences across the two age groups in terms of how they performed relative to chance, showing a systematic preference for the matching face in the female voice condition only in 9-month-old infants. These differences were also reflected in analyses of individual PTLTs.

In particular, 6-month-old infants did not prefer to look at the face that matched the gender of the voice when they heard a female voice, $t(24) = 1.63$, $p = .12$, Cohen's $d = 0.33$, or when they heard a male voice, $t(22) = 0.11$, $p = .92$, Cohen's $d = 0.02$. At the individual level, only 14 of the 25 infants (56%) looked longer at the female face when a female voice was presented, and just 12 of the 23 infants (52%) looked longer at the male face when a male voice was presented (binomial tests, $p = .13$, $p = .16$, respectively). The proportion of matching was not different in the female versus male voice conditions, Fisher exact test, $p = .79$.

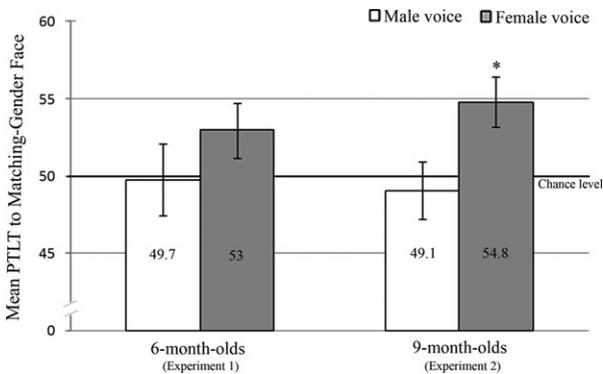


Figure 2 Mean proportion of total looking time directed to the face that matched the heard voice in terms of gender for 6- and 9-month-old infants. *Note:* * $p < .01$.

The results with the 9-month-old infants indicated that they matched female faces and voices, $t(22) = 2.89$, $p < .01$, Cohen's $d = 0.60$, but not male faces and voices, $t(23) = 0.50$, $p = .62$, Cohen's $d = 0.10$. At the individual level, we found that 16 of the 23 infants (70%) looked longer at the female face when a female voice was presented (binomial test, $p < .05$) and that only 10 of 24 infants (42%) looked longer at the male face when a male voice was presented (binomial test, $p = .12$). The proportion of matching in the female voice condition was marginally greater than that in the male voice condition, Fisher exact test, $p = .08$.

To ensure that overall attention could not account for behavioral differences between the two age groups, we analyzed the total amount of looking at the faces with a two-way ANOVA with age (6-month-olds, 9-month-olds) as the between-subjects factor, and the gender of the face (male, female) as the within-subjects factor. This analysis showed that the age effect was not significant, $F(1, 93) = 2.99$, $p = .087$, $\eta_p^2 = 0.03$, suggesting that the difference in audiovisual matching performance at 6 and 9 months of age was not attributable to a difference in overall attention (6-month-olds, $M = 24.5$ sec, $SD = 2.6$ sec; 9-month-olds, $M = 25.1$ sec, $SD = 1.3$ sec). There was a significant effect of the gender of the face, $F(1, 93) = 4.21$, $p < .05$, $\eta_p^2 = 0.04$, whereas the interaction between the gender of the face and age did not reach significance, $F(1, 93) = 0.55$, $p = .46$, $\eta_p^2 = 0.01$. The infants preferred to look at the female faces ($M = 12.9$ sec, $SD = 2.6$ sec) over the male faces ($M = 11.9$ sec, $SD = 2.6$ sec). Follow-up paired two-sample t -tests critically revealed that infants looked more at the female face ($M = 13.5$ sec, $SD = 2.4$ sec) than the male face ($M = 11.5$ sec, $SD = 2.1$ sec) when they heard the female voice, $t(47) = 3.16$, $p < .01$, Cohen's $d = 0.86$, but equally at the female face ($M = 12.2$ sec, $SD = 2.7$ sec) and the male face ($M = 12.2$ sec, $SD = 3$ sec) when they heard the male voice $t(46) = 0.04$, $p = .49$ Cohen's $d = 0.001$.

DISCUSSION

The current study investigated infant perception of multisensory gender coherence. Results showed that 6-month-old infants did not match the audible and visible attributes of gender, but that 9-month-old infants did match these attributes, although only when presented with female faces and voices. These effects were observed in the comparisons of infant performance relative to chance; we did not obtain an age \times voice gender interaction in the ANOVA. Nonetheless, the effects were present in both the group and individual data in the comparisons to chance. Given these

results, it is possible that the absence of the interaction in the ANOVA indicates that the age-related behavioral difference is due to some quantitative (rather than qualitative) change where intersensory matching in older infants is more systematic and less variable than matching in younger infants (see Patterson & Werker, 2002; for similar results in 4.5- to 8-month-olds, and Poulin-Dubois et al., 1994, for similar results in 9- to 12-month-olds). It is also worth noting that whatever degree of developmental change is occurring in the perception of multisensory gender coherence between 6 and 9 months remains incomplete, as older infants only matched the audible and visible attributes of gender for females.

The finding that perception of multisensory gender coherence may emerge between 6 and 9 months of age takes on significance when one considers that young infants can already distinguish and categorize voices (Jusczyk et al., 1992; Miller, 1983; Miller et al., 1982) and faces (Cornell, 1974; Fagan & Singer, 1979; Leinbach & Fagot, 1993; Quinn et al., 2002) in terms of gender. The later emergence of the ability to perceive multisensory gender coherence is consistent with the idea that multisensory perceptual mechanisms develop independently from unisensory perceptual mechanisms even though the former obviously depend on the latter.

At first blush, it might seem that our findings are inconsistent with the Walker-Andrews et al. (1991) result that matching of the auditory and visual attributes of gender occurs as early as 6 months of age. A closer examination of the specific procedures used in the Walker-Andrews et al. study reveals, however, that the gender of the voice was changed between the first and the second half of the experiment. This type of change highlighted the audible gender difference and, by doing so, probably facilitated the ability of younger infants to choose the matching face. In contrast, in the present study, as well as in other previous studies where gender did not vary across trials, only older infants exhibited evidence of audiovisual matching (Patterson & Werker, 2002; Poulin-Dubois et al., 1994).

In addition, as previously mentioned, a crucial issue in the Walker-Andrews et al. study is that the voices heard by the infants were the actual voices of the two speakers whose faces were shown. This aspect of the procedure raises the possibility that idiosyncratic information might have facilitated matching because speakers differ in the way they pronounce individual segmental and suprasegmental information (Dohen, Løevenbrück, & Hill, 2009; Johnson, Ladefoged, & Lindau, 1993; Keating, Lindblom, Lubker, & Kreiman, 1994). Because only two people were presented in Walker-Andrews et al. and because these people had been instructed to speak in synchrony during the recordings, they may each have produced idiosyncratic articulatory or respiratory patterns. Therefore, the possibility that idiosyncratic properties of the voices and faces

facilitated matching in the Walker-Andrews et al. study cannot be ruled out. In contrast, in our study, for each of the face pairs presented, we used a voice that did not correspond to any of the faces. This procedure ensured that the matching could not be based on idiosyncratic properties.

Like in previous studies of older infants, we found that infants only matched the gender of female faces and voices. The most likely explanation for this asymmetrical responsiveness to female versus male faces and voices is that infants typically have more experience with females as opposed to males (Rennels & Davis, 2008; Sugden, Mohamed-Ali, & Moulson, 2014). Also of note is that the late emergence of the ability to perceive multisensory gender coherence contrasts with the earlier emergence of the ability to perceive the multisensory coherence of other attributes. The fact is, however, that the detection of face and voice gender relations requires previous learning of arbitrary multisensory associations (e.g., low pitch and prominent eyebrows). Even though infants as young as 3 months of age already exhibit the ability to associate individual voice characteristics with particular faces (Brookes et al., 2001; Spelke & Owsley, 1979; Trehub et al., 2009), the present results suggest that infants require more extended experience with the perceptual characteristics of gender before they can recognize its multisensory coherence. The later emergence of multisensory gender categories (i.e., male versus female) relative to multisensory identity categories (e.g., Harry versus Sally) may reflect the fact that each gender category would need to be maintained across a range of variation of values along each of the relevant attributes (e.g., pitch for voices, eyebrow prominence for faces) for the different exemplars of the category.

Finally, it should be noted that task complexity plays a role in infant perception of multisensory gender coherence. Put differently, the detection of multisensory gender coherence is not an all-or-none phenomenon, but rather depends on what is required of the infants. This principle is illustrated by findings showing that, on the one hand, infants as young as four months of age can perceive multisensory gender coherence when it is specified by the visible and audible attributes of an isolated vowel (Patterson & Werker, 2002). When, on the other hand, infants have to extract and match audible and visible gender information from continuous streams of auditory and visual information, they appear unable to do so until several months later. The fact that task complexity must be taken into account indicates that our initial hypothesis that prosodic variation of audiovisual gender information would make it easier for infants to perceive the multisensory coherence of gender was too simplistic. Rather, it appears that infant multisensory responsiveness depends on a combination of factors, including stimulus complexity, task complexity, and the specific degree of

infant perceptual expertise. This more nuanced theoretical perspective is probably a more accurate description of the mechanisms underlying the development of multisensory perception and suggests new avenues for future research. For example, future studies might contrast perception of multisensory gender coherence in infants raised primarily by female caregivers versus those raised primarily by male caregivers. Presumably, this sort of differential experience with a specific gender category would have asymmetrical effects on infant ability to perceive female versus male multisensory gender coherence.

ACKNOWLEDGMENTS

This work was supported by Grant R01 HD-46526 from the National Institute of Child Health and Human Development. We thank two anonymous reviewers for their comments.

REFERENCES

- Barenholtz, E., Lewkowicz, D. J., Davidson, M., & Mavica, L. (2014). Categorical congruence facilitates multisensory associative learning. *Psychonomic Bulletin & Review*, *21*, 1346–1352.
- Brookes, H., Slater, A., Quinn, P. C., Lewkowicz, D. J., Hayes, R., & Brown, E. (2001). Three-month-old infants learn arbitrary auditory-visual pairings between voices and faces. *Infant and Child Development*, *10*, 75–82.
- Brown, E., & Perrett, D. I. (1993). What gives a face its gender? *Perception*, *22*, 829–840.
- Coleman, R. O. (1971). Male and female voice quality and its relationship to vowel formant frequencies. *Journal of Speech and Hearing Research*, *14*, 565–577.
- Cornell, E. H. (1974). Infants' discrimination of photographs of faces following redundant presentations. *Journal of Experimental Child Psychology*, *18*, 98–106.
- Coulon, M., Guellai, B., & Streri, A. (2011). Recognition of unfamiliar talking faces at birth. *International Journal of Behavioral Development*, *35*, 282–287.
- Dohen, M., Lœvenbruck, H., & Hill, H. (2009). Recognizing prosody from the lips: Is it possible to extract prosodic focus from lip features? In A. Wee-Chung Liew & S. Wang (Eds.), *Visual speech recognition: Lip segmentation and mapping* (pp. 416–438). Hershey, PA: Medical Information Science Reference.
- Fagan, J. F., & Singer, L. T. (1979). The role of simple feature differences in infants' recognition of faces. *Infant Behavior and Development*, *2*, 39–45.
- Fernald, A. (1985). Five-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*, 181–195.
- Gelfer, M. P., & Mikos, V. A. (2005). The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. *Journal of Voice*, *19*, 544–554.
- Guellai, B., Coulon, M., & Streri, A. (2011). The role of motion and speech in face recognition at birth. *Visual Cognition*, *19*, 1212–1233.

- Johnson, K., Ladefoged, P., & Lindau, M. (1993). Individual differences in vowel production. *Journal of the Acoustical Society of America*, *94*, 701–714.
- Jusczyk, P. W., Pisoni, D. B., & Mullennix, J. (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition*, *43*, 253–291.
- Kamachi, M., Hill, H., Lander, K., & Vatikiotis-Bateson, E. (2003). 'Putting the face to the voice': Matching identity across modality. *Current Biology*, *13*, 1709–1714.
- Keating, P. A., Lindblom, B., Lubker, J., & Kreiman, J. (1994). Variability in jaw height for segments in English and Swedish VCVs. *Journal of Phonetics*, *22*, 407–422.
- Kim, H. I., & Johnson, S. P. (2014). Detecting 'infant-directedness' in face and voice. *Developmental Science*, *17*, 621–627.
- Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, *87*, 820–857.
- Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*, 1138–1141.
- Ladefoged, P., & Broadbent, D. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, *29*, 98–104.
- Leinbach, M. D., & Fagot, B. I. (1993). Categorical habituation to male and female faces: Gender schematic processing in infancy. *Infant Behavior and Development*, *16*, 317–332.
- Lewkowicz, D. J., Minar, N. J., Tift, A. H., & Brandon, M. (2015). Perception of the multi-sensory coherence of fluent audiovisual speech in infancy: Its emergence and the role of experience. *Journal of Experimental Child Psychology*, *130*, 147–162.
- Miller, C. L. (1983). Developmental changes in male/female voice classification by infants. *Infant Behavior and Development*, *6*, 313–330.
- Miller, C. L., Younger, B. A., & Morse, P. A. (1982). The categorization of male and female voices in infancy. *Infant Behavior and Development*, *5*, 143–159.
- Nakata, T., & Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behavior and Development*, *27*, 455–464.
- Noles, E. C., Kayl, A. J., & Rennels, J. L. (2012). *Dynamic presentation does not augment infants' intermodal knowledge of males*. Paper presented at the International Conference on Infant Studies, Minneapolis, USA.
- Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, *22*, 237–247.
- 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.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, *24*, 175–185.
- Poulin-Dubois, D., Serbin, L. A., & Derbyshire, A. (1998). Toddlers' intermodal and verbal knowledge about gender. *Merrill-Palmer Quarterly*, *44*, 338–354.
- Poulin-Dubois, D., Serbin, L. A., Kenyon, B., & Derbyshire, A. (1994). Infants' intermodal knowledge about gender. *Developmental Psychology*, *30*, 436–442.
- 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*, 1109–1122.
- Ramsey, J. L., Langlois, J. H., & Marti, N. C. (2005). Infant categorization of faces: Ladies first. *Developmental Review*, *25*, 212–246.
- Rennels, J. L., & Davis, R. E. (2008). Facial experience during the first year. *Infant Behavior and Development*, *31*, 665–678.

- Rosenblum, L. D., Yakel, D. A., Baseer, N., Panchal, A., Nodarse, B. C., & Niehus, R. P. (2002). Visual speech information for face recognition. *Perception & Psychophysics*, *64*, 220–229.
- Sachs, J., Lieberman, P., & Erickson, D. (1973). Anatomical and cultural determinants of male and female speech. In R. W. Shuy, & R. W. Fasold (Eds.), *Language attitudes* (pp. 74–83). Washington, DC: Georgetown University Press.
- 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.
- Spelke, E. S., & Owsley, C. J. (1979). Intermodal exploration and knowledge in infancy. *Infant Behavior and Development*, *2*, 13–27.
- Sugden, N. A., Mohamed-Ali, M. I., & Moulson, M. C. (2014). I spy with my little eye: Typical, daily exposure to faces documented from a first-person infant perspective. *Developmental Psychobiology*, *56*, 249–261.
- Trehub, S. E., Plantinga, J., & Brcic, J. (2009). Infants detect cross-modal cues to identity in speech and singing. *Annals of the New York Academy of Sciences*, *1169*, 508–511.
- Van Borsel, J., Janssens, J., & De Bodt, M. (2009). Breathiness as a feminine voice characteristic: a perceptual approach. *Journal of Voice*, *23*, 291–294.
- Walker-Andrews, A. S., Bahrick, L. E., Raglioni, S. S., & Diaz, I. (1991). Infants' bimodal perception of gender. *Ecological Psychology*, *3*, 55–75.
- Xiao, N. G., Perrotta, S., Quinn, P. C., Wang, Z., Sun, Y.-H. P., & Lee, K. (2014). On the facilitative effects of face motion on face recognition and its development. *Frontiers in Psychology*, *5*, 633.
- Younger, B. A., & Fearing, D. D. (1999). Parsing items into separate categories: Developmental change in infant categorization. *Child Development*, *70*, 291–303.