THE OFFICIAL JOURNAL OF THE
INTERNATIONAL CONGRESS
OF INFANT STUDIES

Infancy, **24**(4), 570–588, 2019

ISSN: 1525-0008 print / 1532-7078 online
DOI: 10.1111/infa.12290

WILEY Blackwell

Resolving the (Apparent) Talker Recognition Paradox in Developmental Speech Perception

Natalie Fecher
Department of Psychology
University of Toronto Mississauga

Melissa Paquette-Smith
Department of Psychology
University of California, Los Angeles

Elizabeth K. Johnson
Department of Psychology
University of Toronto Mississauga

The infant literature suggests that humans enter the world with impressive built-in talker processing abilities. For example, newborns prefer the sound of their mother's voice over the sound of another woman's voice, and well before their first birthday, infants tune in to language-specific speech cues for distinguishing between unfamiliar talkers. The early childhood literature, however, suggests that preschoolers are unable to learn to identify the voices of two unfamiliar talkers unless these voices are highly distinct from one another, and that adult-level talker recognition does not emerge until children near adolescence. How can we reconcile these apparently paradoxical messages conveyed by the infant and early childhood literatures? Here, we address this question by testing 16.5-month-old infants ($N = 80$) in three talker recognition experiments. Our results demonstrate that infants at this age have difficulty recognizing unfamiliar talkers, suggesting that talker recognition (associating voices with people) is mastered later in life than talker discrimination (telling voices apart). We conclude that methodological differences across the infant and early childhood literatures—rather than a true developmental discontinuity—account for the performance differences in talker processing between these two age groups. Related findings in other areas of developmental psychology are discussed.

Soon after the auditory system becomes operational in the third trimester of pregnancy, humans demonstrate surprisingly sophisticated talker processing abilities. Fetuses react differently to their mother's voice than the voice of a female stranger

Correspondence should be sent to Elizabeth K. Johnson, Department of Psychology, University of Toronto Mississauga, 3359 Mississauga Road, Mississauga, ON L5L 1C6, Canada. E-mail: elizabeth.johnson@utoronto.ca

(e.g., Kisilevsky & Hains, 2011), newborns preferentially listen to their mother's voice (e.g., DeCasper & Fifer, 1980), and 7.5-month-olds readily tell apart unfamiliar female voices (e.g., Fecher & Johnson, 2018b). Indeed, infants seem to be so attuned to indexical (i.e., talker-specific) detail in speech that they sometimes fail to recognize a previously familiarized word form when it is produced by a novel talker (e.g., Houston & Jusczyk, 2000). Yet, much like the development of face recognition, children do not develop adult-like talker recognition abilities until their school-age years (e.g., Mann, Diamond, & Carey, 1979). For example, 3- to 6-year-olds are much less accurate than adults at learning to identify two unfamiliar talkers (Creel & Jiménez, 2012), and 5- and 6-year-olds are significantly worse than adults at recognizing four unfamiliar female talkers in a "voice line-up" procedure (Fecher & Johnson, 2018a). Some studies even suggest that children do not attain adult-level talker processing skills until adolescence (e.g., Levi & Schwartz, 2013; Mann et al., 1979). These findings raise the question: Why does the infant literature suggest that humans enter the world with excellent talker processing skills, whereas the early childhood literature focuses more on children's *poor* performance in this area relative to adults? Could differences in the way we experimentally assess talker processing in infants versus children explain these perceived—and seemingly paradoxical—differences in talker processing across ages? Here, we address these questions by testing 16.5-month-old infants on a talker recognition task. Our findings lead us to conclude that although infants are prepared to identify the most socially relevant talkers in their environment (e.g., their mothers), the development of mature talker recognition abilities takes time and experience.

Since early work on talker identity processing in infants in the 1970s (e.g., Mills & Melhuish, 1974; Turnure, 1971), it has been established that fetuses, newborns, and infants can tell apart their mother's voice and an unfamiliar female voice, and that they prefer to listen to their mother's voice over the voice of a female stranger. For example, fetuses show significant changes in heart rate in response to hearing their mother's voice but not in response to hearing another woman's voice (e.g., Hepper, Scott, & Shahidullah, 1993; Kisilevsky & Hains, 2011; Kisilevsky et al., 2009; Voegtline, Costigan, Pater, & DiPietro, 2013). Using the non-nutritive sucking procedure, neonates will suck preferentially to hear their mother's voice over an unfamiliar female voice (e.g., DeCasper & Fifer, 1980; Mehler, Bertoncini, Barrière, & Jassik-Gerschenfeld, 1978; Mills & Melhuish, 1974). Newborns and young infants do not show preferences for their father's voice over an unfamiliar male voice, but they can distinguish between the two (e.g., DeCasper & Prescott, 1984; Ward & Cooper, 1999). Furthermore, investigations into the neurophysiological mechanisms underlying the maternal voice preference have shown that the infant brain responds differently to familiar and unfamiliar voices (e.g., Beauchemin et al., 2011; Purhonen, Kilpeläinen-Lees, Valkonen-Korhonen, Karhu, & Lehtonen, 2004).

But can infants tell apart the voices of people they do not know? Findings from adult research on familiar versus unfamiliar talker processing (e.g., Stevenage, 2017; Van Lancker & Kreiman, 1987) suggest that the results of infant studies on maternal voice processing may not generalize to processing of unfamiliar voices (i.e., voices that infants are hearing for the first time rather than voices of close family members). However, few studies have put infants' ability to distinguish unfamiliar voices to the test. In one study, researchers habituated near-term fetuses to either a male voice or a female voice and subsequently tested fetal cardiac responses to a voice change (Lecanuet, Granier-Deferre, Jacquet, Capponi, & Ledru, 1993). Fetuses responded with a significant change in heart rate when hearing the new voice (which always differed in gender from the habituated voice). Moreover, neonates can distinguish between female tokens of a disyllabic word and male tokens of the same word (Floccia, Nazzi, & Bertoncini, 2000), and 6- and 7-month-old infants can categorize voices based on gender (e.g., Miller, 1983; Miller, Younger, & Morse, 1982). Taken together, these findings suggest that infants can tell apart male and female talkers very early on.

Testing infants on voices that differ in gender may, however, lead us to overestimate infants' voice processing skills, because men's and women's voices are typically quite distinct from one another and therefore relatively easy to tell apart. In addition, the task of telling apart different-gender voices does not adequately reflect the challenges that infants face in their everyday listening environment, in which they are often exposed to multiple talkers of the same gender. To date, only three infant studies have examined the processing of unfamiliar same-gender talkers. In two of these studies, 7.5-month-olds were habituated to multiple female talkers and then tested on whether they could distinguish the habituated talkers from a novel female talker (Fecher & Johnson, 2018b; Johnson, Westrek, Nazzi, & Cutler, 2011). Infants could successfully tell the talkers apart provided that the talkers produced the test sentences in the infants' native language. In keeping with this finding, a study using the conditioned headturn procedure showed that 6- and 12-month-olds could discriminate between two female talkers who each produced several tokens of a bisyllabic word (Friendly, Rendall, & Trainor, 2014). Hence, infants seem to be capable of telling apart even same-gender talkers from a very young age (at least female talkers).

The picture that emerges is that infants appear to be quite proficient at distinguishing between talkers even when the talkers are unfamiliar and matched in gender. We will henceforth refer to the ability to tell talkers apart (which can be accomplished by attending to low-level acoustic-phonetic information) as *talker discrimination*. But while past studies have taught us a lot about infants' talker discrimination skills, no study has previously tested unfamiliar *talker recognition* in children under age 3 (see Bahrick, Hernandez-Reif, & Flom, 2005, for an intermodal matching study). That is, so far we know virtually nothing about infants' ability to learn to associate unfamiliar voices with specific people (that are not Mom or Dad) or to recall voices after longer periods of time (i.e., beyond the time needed to make a "same" or "different" judgment from one experimental trial to the next).

Research on unfamiliar talker recognition (rather than discrimination) in infants would not only extend our knowledge of indexical processing in infancy, it might also explain why the infant and early childhood literatures paint what could be perceived as contradictory pictures of infants' and children's talker processing skills. Here, we speculate that what on the surface could be seen as a developmental paradox in talker recognition can be explained by methodological differences across the two literatures. In the past, infants were almost exclusively tested on talker discrimination tasks, while older children were largely tested on talker recognition tasks, including talker learning tasks (e.g., Creel & Jiménez, 2012; Perea et al., 2014) and voice line-ups (e.g., Fecher & Johnson, 2018a; Öhman, Eriksson, & Granhag, 2011). Discrimination and recognition tasks differ in many ways, and they put different demands on the auditory system (as previously discussed in, e.g., Creel & Jiménez, 2012; Fecher & Johnson, 2018a; Sadakata & McQueen, 2013; Werker, Fennell, Corcoran, & Stager, 2002; Winters, Levi, & Pisoni, 2008). Given these discrepancies across the two literatures, it is perhaps not too surprising that the developmental time course for talker identity processing is difficult to map out. In the current study, we used a task that is more comparable to the tasks typically used with older children, allowing us to test infants' talker recognition (rather than discrimination) skills. We tested infants between 16 and 17 months of age because we wanted our participants to be considerably younger than the children tested in previous talker recognition studies. At the same time, we needed to ensure that our task—which involved learning two face-voice pairings at once—was suitable for testing this age group. Testing 16.5-month-olds seemed appropriate because past work has shown that younger infants sometimes struggle with learning two novel audiovisual pairings (e.g., 20-month-olds but not 14-month-olds can learn two novel word-object pairings at once; Werker et al., 2002).

To summarize, the current study tested the hypothesis that the differences in talker processing between infants and children arise from differences in the task demands associated with the paradigms that are typically used with these age groups. In Experiment 1, we tested infants' ability to recognize voices paired with animated cartoon characters. In Experiments 2 and 3, we tested infants' ability to recognize the voices of talking human faces. We conclude that talker recognition is a complex and cognitively challenging task for infants, and that the developmental paradox in talker recognition can be resolved by considering the task differences across the infant and early childhood literatures.

EXPERIMENT 1

In Experiment 1, we used an infant-friendly talker recognition task to test whether 16.5-month-olds can learn to recognize the voices of animated cartoon characters after brief exposure to two voice-character pairings. Using cartoons allowed us to control for the visual similarity of the characters and thus to reduce the likelihood that infants would show preferences for one character over the other. During training, infants saw two characters take turns speaking. During test, infants saw still images of both characters side-by-side and heard the voice of one of them. If infants recognized which character was talking, then they should look toward the appropriate character (i.e., their looking time to the character that matched the voice should be longer than their looking time to the character that did not match the voice). Hence, we did not test whether infants could tell the voices apart but whether they could recognize who was talking (i.e., remember which voice belonged to which character). One group of infants was exposed to pairs of female talkers during training and test (same-gender condition), and a second group of infants was presented with pairs of one female and one male talker (mixed-gender condition). For the mixed-gender condition, we predicted that infants would successfully recognize the voices (because male and female voices are typically quite distinct from one another). For the same-gender condition, our prediction was less straightforward. Based on the infant literature, which suggests that infants are quite proficient at talker discrimination, we would predict that infants succeed in the talker recognition task (even with all-female voices). Based on the early childhood literature, which implies that even teenagers have not yet reached mature levels of talker recognition, we would predict that infants perform poorly in this task. If our original hypothesis were correct, and the apparent developmental differences in talker processing are in fact caused by task differences, then the latter outcome would seem more likely.

Method

Participants

Thirty-two English-learning 16- to 17-month-old infants (M_{age} = 502 days, range = 485–518; 12 female) from the Greater Toronto Area were tested. No hearing or vision difficulties and no recent ear infections were reported. All infants were exposed to English at least 90% of the time. The data for three additional infants (one in the mixed-gender and two in the same-gender condition) were excluded prior to data analysis due to fussing. For all three reported experiments, infants' caregivers provided informed written consent for their child's participation. This study was approved by the University of Toronto Research Ethics Board.

Design

Half of the infants were randomly assigned to the same-gender condition, and the other half were assigned to the mixed-gender condition. The use of training and test passages, the screen that the characters appeared on, and the order of presentation of trials were counterbalanced across infants. The assignment of characters to talkers was also counterbalanced. For example, half of the infants saw a green bear paired with the female talker and a purple bear paired with the male talker, and the other half saw a green bear paired with the male talker and a purple bear paired with the female talker.

Stimuli

Native Canadian English-speaking male and female talkers (M_{age} = 22.3 years, *SD* = 5.9) were recorded reading the 10 English passages shown in Appendix A (drawn from Paquette-Smith & Johnson, 2016; 13–17 syllables per passage). The talkers were instructed to read the passages in a child-directed manner with a "happy" tone of voice. Recordings were made in a double-walled, sound-attenuated Industrial Acoustics Company (IAC) booth (44.1 kHz; normalized to 69.5 dB). Additional information about how the talkers were paired across conditions, and basic acoustic measurements of the talkers' speech productions, are provided in Table A1 in Appendix B.

Visual stimuli consisted of animated cartoon bears and bunnies that were created using software which allows users to produce short video clips by synchronizing the mouth movements of a cartoon character with recorded speech (Zoobe Message Entertainment GmbH, Berlin, Germany). All characters used in our study had gender-neutral appearances and were matched in size and brightness. The animations were visually appealing to young children (e.g., characters were jumping, spinning, and waving). Per trial, we presented infants either with a pair of purple and green bears or with a pair of orange and turquoise bunnies to reduce the possibility that infants would prefer to look at one animal over the other.

Procedure

Participants were tested using a variant of the preferential looking procedure. Infants sat on their caregiver's lap in an IAC booth and faced two side-by-side 21.5-

inch computer screens. A high-definition camcorder located below the two screens was used to record infants' eye movements for offline coding. High-quality loudspeakers presented the audio at a constant, comfortable volume. Caregivers wore noise-canceling headphones and listened to masking music intermixed with stimuli used in the experiment to prevent them from influencing their child's performance.

Each infant completed two experimental blocks, with each block consisting of a training phase immediately followed by a test phase (see Figure 1). In each block, infants were trained and tested on two voices (four voices in total). During each of six training trials per block, infants saw one of the two characters speaking (two passages per trial). Each character consistently appeared in the center of either the left or right screen (i.e., the side that the characters appeared on never changed per infant), which increased the likelihood that infants would succeed in the task. While one of the characters was animated and speaking, a still image of the other character was shown on the opposite screen. Speech started 200 msec after trial onset and speech offset was followed by at least 500 msec of acoustic silence before the end of the trial. The final silence period differed slightly across trials because we ended trials only once the speaking characters had completed their current motion (e.g., spinning in a circle) and had returned to their starting position (*M* trial length = 10.1 sec, *SD* = 0.8). Each training phase was followed by two test trials, where both characters appeared motionless on their appointed screen and infants heard one of the voices from the training phase (one passage per trial; different passages were used for training and test). Speech again started 200 msec after trial onset and was followed by 500 msec of silence before the end of the trial (*M* trial length = 5.1 sec, *SD* = 0.7). We predicted that if infants recognized the voice, then they should look toward the character that was paired with the voice during training. Between trials, a 2-sec flashing blue star was used to attract

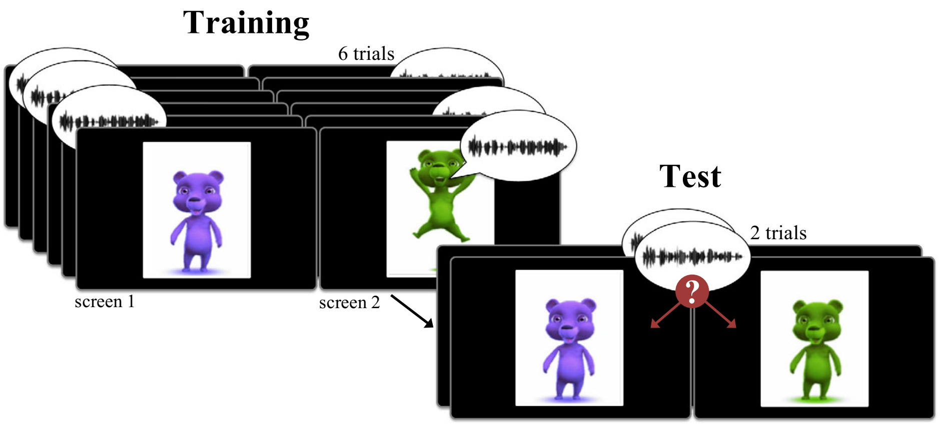

Figure 1 Sample experimental block in Experiment 1. Each six-trial training phase was followed by a two-trial test phase. During training, infants were familiarized with two cartoon characters (which differed in color but otherwise looked identical). The speaking character was animated while the silent character stood still. During test, infants saw still images of both characters and heard one of the voices from the training phase. If infants recognized the voice, then they should look toward the appropriate character. In the second experimental block, infants were trained and tested on a new pair of voices and characters. Infants assigned to the same-gender condition were presented with pairs of female voices in both blocks, while infants assigned to the mixed-gender condition were presented with pairs of one female and one male voice.

infants' attention back to the center of the two screens. Between blocks, infants watched a 7-sec cartoon (without speech) to keep their attention and clearly separate the two blocks. The experiment took approximately 3.1 min to complete.

Coding and analysis

All coding was done using SuperCoder 1.5 (Hollich, 2005) with the audio track disabled. We coded each 33 msec frame as a look to the left or right character (or neither). Four randomly-selected videos were recoded by a second coder, and inter-coder reliability was high (Mean $r = .96$, $SD = .02$). The analysis focused on the portion of the trial during which the character was speaking. The window of analysis started 1 sec after speech onset (1.2 sec after trial onset because of the 200 msec silence before speech onset). We started the window of analysis 1 sec after speech onset because infants needed time to program an eye movement and to potentially orient their head toward one of the two adjacent screens upon hearing the voice. In addition, based on what we know about adult voice recognition, we speculated that infants need to hear at least 2–3 syllables of speech before they have a fair chance at recognizing who is talking (e.g., Bricker & Pruzansky, 1966; Cook & Wilding, 1997; Yarmey & Matthys, 1992). Although the talkers spoke at slightly different speech rates, we wanted to use the same analysis window for all trials. Thus, the window of analysis was determined by the duration of the shortest test passage (resulting in a 2.5 sec window).

Results and discussion

To assess infants' voice recognition abilities, we analyzed the average proportion of looks to target in the 2.5 sec window of analysis starting 1 sec after speech onset. If infants recognized which character was talking, then they should look significantly longer at the image of that character than would be expected by chance (.5).

A one-way analysis of variance (ANOVA) with the proportion of looks to target as the dependent variable and condition (same-gender, mixed-gender) as an independent variable revealed a significant main effect of condition, $F(1, 30) = 4.35$, $p = .046$, $\eta_p^2 = .13$. This suggests that infants' looking behavior significantly differed between the same- and mixed-gender conditions (see Figure 2, left panel). In the same-gender condition, infants performed at chance ($M = .48$, $SD = .15$), $t(15) = -0.57$, $p = .576$, indicating that they did not look longer toward the appropriate character when hearing the voice. In the mixed-gender condition, however, the proportion of looks to target was significantly above chance ($M = .59$, $SD = .15$), $t(15) = 2.36$, $p = .032$, $d = 0.59$, suggesting that infants recognized who was talking.

In summary, Experiment 1 showed that infants could learn to recognize the voices of two animated cartoon characters when the talkers differed in gender. However, infants failed to recognize the talkers when they were acoustically less distinct from one another (all-female talkers). The finding that infants performed significantly above chance in the mixed-gender condition suggests that our task was appropriate for testing talker recognition in 16.5-month-olds. Nevertheless, infants' performance was not overwhelmingly strong (even in the mixed-gender condition) thus raising the question: Why did infants perform so poorly?

One explanation for infants' overall low performance concerns the ecological validity of Experiment 1. It is possible that the task was unnatural for children at this age

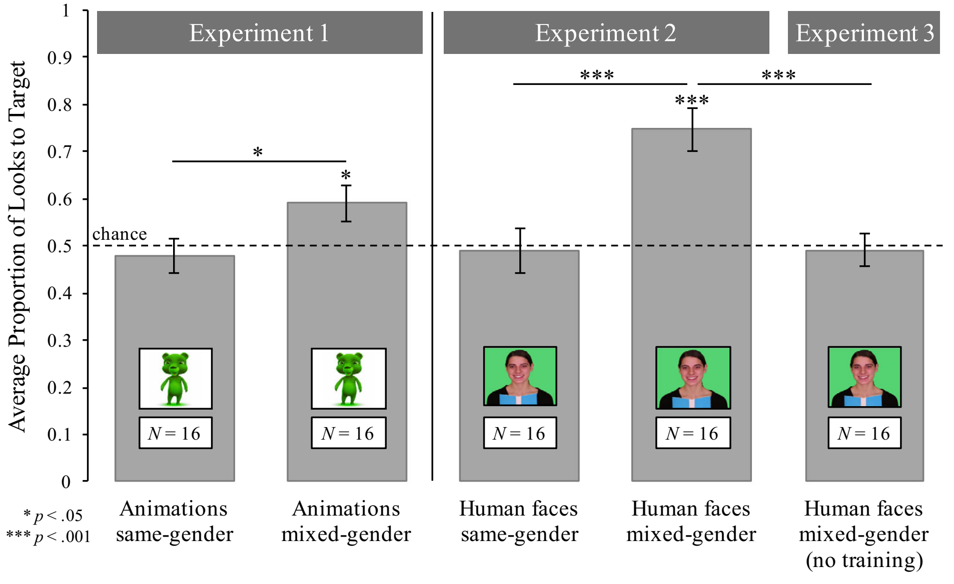

Figure 2 Average proportion of looks to target in Experiment 1 (animated cartoon characters) and Experiments 2 and 3 (talking human faces). In the same-gender conditions in Experiments 1 and 2 (all-female talker pairs), infants did not recognize the talkers (as evidenced by their chance-level performance). However, in the mixed-gender conditions in Experiments 1 and 2 (male/female talker pairs), infants recognized the talkers. In Experiment 3 (no familiarization with the talkers prior to test), infants failed to match the faces and voices based on gender information alone. Error bars represent standard error of the mean.

and lacked an important social component that would have increased infants' attention and motivation to learn the voice-character pairings (see also Creel & Jiménez, 2012). In addition, 16.5-month-olds typically have more experience with integrating auditory and visual information in talking human faces than they have with relating human voices to computer animations (e.g., Bahrick et al., 2005; Brookes et al., 2001; Burnham & Dodd, 2004; Patterson & Werker, 2003; Trehub, Plantinga, & Brcic, 2009; Walker-Andrews, 1997; Walker-Andrews, Bahrick, Raglioni, & Diaz, 1991). We therefore hypothesized that infants might be better at recognizing voices if they learned the voices from talking human faces rather than animated cartoon characters. In Experiment 2, we put this hypothesis to the test.

EXPERIMENT 2

In Experiment 1, infants had difficulty learning to recognize previously unfamiliar voices when the voice referents were visibly not human. Although infants performed above chance in the mixed-gender condition, their performance was only at 59% (with chance performance being 50%); and in the same-gender condition, infants failed to recognize the talkers altogether. Based on the current data and earlier findings on audiovisual speech perception in infancy, we hypothesized that testing infants on talking human faces instead of cartoons might facilitate talker recognition. In Experiment

2, we tested a new group of 16.5-month-olds on the faces and voices of male and female talkers. We predicted that these infants would be more successful at talker recognition than the infants tested in Experiment 1, and that they would again perform better in the mixed-gender than same-gender condition.

Method

Participants

Thirty-two English-learning 16- to 17-month-olds (M_{age} = 506 days, range = 490–520; 19 female) from the Greater Toronto Area were tested (eligibility criteria were the same as in Experiment 1). The data for eight additional infants (one in the mixed-gender and seven in the same-gender condition) were excluded due to fussing.

Design

As in Experiment 1, half of the infants were randomly assigned to the same-gender condition and the other half were assigned to the mixed-gender condition. The use of training and test passages, the screen that the talkers appeared on, and the order of presentation of trials, were counterbalanced across infants.

Stimuli

Native Canadian English-speaking male and female talkers (M_{age} = 23.9 years, SD = 3.2) produced the 10 training passages from Experiment 1 and the four test passages shown in Appendix A (10 syllables per passage; see Table A1 in Appendix B for additional talker characteristics). All test passages ended with the phrase "Look at me!" to more explicitly prompt infants to look at the talker that matched the voice. Critically, the talkers were well matched in visual appearance (e.g., skin and hair color, hairstyle, attractiveness) and emotional affect (i.e., how friendly they looked and sounded). This was again aimed at compensating for potential preferences for one talker over the other (see, e.g., Langlois, Roggman, Casey, Ritter, & Rieser-Danner, 1987).

Recordings were made in a sound-attenuated recording studio. Talkers were seated in front of a plain green background, and the camera recording them was positioned so that the images consisted of the talkers' head and shoulders (see Figure 3). All talkers wore black T-shirts, all women wore their hair in a ponytail, and glasses and jewelry were removed.

Procedure

Previous studies have used an intermodal matching procedure to examine infants' ability to detect a mismatch between previously familiarized faces and voices (e.g., Bahrick et al., 2005; Brookes et al., 2001). However, it could be argued that this task lacks ecological validity and does not necessarily reflect the rapid talker recognition abilities required in real-world social situations. Since the goal of our study was to examine whether infants can learn to associate a voice with a particular person, we instead used a variant of the preferential looking procedure (as in Experiment 1).

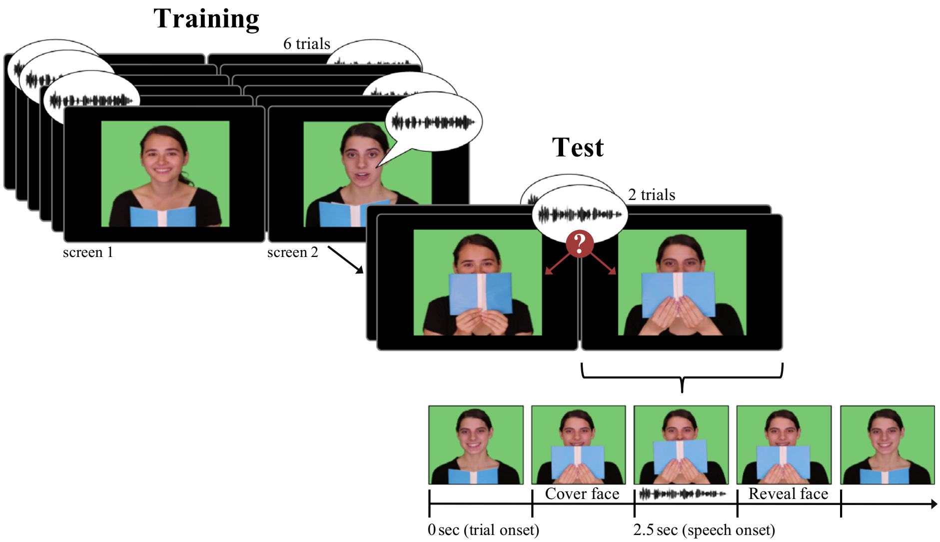

Figure 3 Sample experimental block in Experiment 2. Each six-trial training phase was followed by a two-trial test phase. During training, infants were familiarized with two talkers. While one talker spoke, the other talker was shown smiling at the camera. During test, infants saw both talkers with the lower part of their face concealed and heard the voice of one of the talkers. If infants recognized the voice, then they should look toward the person that matched the voice. In the second experimental block, infants were trained and tested on a different pair of talkers. Infants assigned to the same-gender condition were exposed to pairs of female talkers in both blocks, and infants assigned to the mixed-gender condition were presented with pairs of one female and one male talker.

Infants were familiarized with the faces and voices of two talkers, and we subsequently tested whether infants would look toward the correct talker upon hearing the voice. A critical element of our design was to prevent infants from expecting to see the talker's mouth moving while hearing the speech. To achieve this goal, we could have simply shown still images of the talkers. However, to maximize ecological validity, we presented infants with videos of the talkers in which each talker's mouth region was occluded by a book. This way the voice could plausibly be associated with either talker. Importantly, while infants listened to the test passages, they saw videos of the talkers when the talkers were in fact *not* speaking. This ensured that infants relied on vocal cues to talker identity rather than visual speech cues (e.g., Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004).

Infants completed two experimental blocks, with each block consisting of a training phase and a test phase (see Figure 3). During each of six training trials per block, infants watched one of the two talkers reading to them while holding a small blue book at chest height (two training passages per trial). Each talker consistently appeared in the center of either the left or the right screen (i.e., as in Experiment 1, the talkers never switched sides per infant). While one talker spoke, the other talker was shown smiling at the camera on the opposite screen (M trial length = 9.6 sec, SD = 0.7). During each of two test trials per block, the talkers appeared on their appointed screen (with their entire face visible). They then each moved a book up to cover the lower part of their face (up to the nose; see Figure 3). At 2.5 sec after trial onset (once the motion was completed and the book occluded each talker's mouth region), infants heard the voice of one of the talkers (two repetitions of one test passage per trial). We predicted that if infants recognized the voice, then they should look at the appropriate talker. It was not until the passages had finished playing (and no later than 10 sec into the trial) that both talkers moved the book back down (*M* trial length = 12.4 sec, *SD* = 0.2). A 2-sec flashing blue star served to bring the infant's attention back to the center of the screens at the end of each trial, and a short cartoon was shown between blocks to separate the blocks and keep infants engaged. The experiment lasted for approximately 3.5 min.

Coding and analysis

Videos were coded in the same way as that reported for Experiment 1. Reliability between two independent coders was high (Mean r = .97, *SD* = .04). The analysis focused on the portion of the trial during which the talker was speaking, which coincided with the portion during which each talker's face was partially covered by the book. The window of analysis was chosen to be the same as the window used in Experiment 1; that is, it lasted for 2.5 sec and started 1 sec after speech onset (3.5 sec after trial onset because of the 2.5 sec silence prior to speech onset, which was the time needed to cover the face with the book).

Results and discussion

As in Experiment 1, the looking time data from Experiment 2 were analyzed using a 2.5 sec window of analysis starting 1 sec after speech onset. If infants recognized the voice, then their average proportion of looks to the appropriate talker should be significantly greater than chance level (.5).

A one-way ANOVA revealed a significant main effect of condition (same-gender, mixed-gender) on the proportion of looks to target, $F(1, 30) = 15.91$, $p < .001$, $\eta_p^2 = .35$, suggesting that infants' performance was again dependent on whether they were tested on all-female or male/female talker pairs (see Figure 2, right panel, two leftmost bars). In the same-gender condition, the proportion of looks to target did not significantly differ from chance (*M* = .49, *SD* = .19), $t(15) = -0.21$, $p = .839$; that is, infants did not recognize the voices when all talkers were female. In the mixed-gender condition, the proportion of looks to target was significantly above chance (*M* = .75, *SD* = .18), $t(15) = 5.58$, $p < .001$, $d = 1.40$, showing that infants remembered which voice belonged to which talker when the talkers differed in gender.

Consistent with Experiment 1, infants in Experiment 2 successfully recognized the voices at test when the voices differed in gender. But even with this more ecologically valid testing procedure, infants failed in the same-gender condition of this task. Design differences between Experiments 1 and 2 do not allow us to directly compare findings across experiments, but descriptively speaking, our results for the mixed-gender condition could be taken as an indication that voice learning is more efficient when human talkers are presented (see Figure 2 and larger effect sizes in Experiment 2). However, this finding could also be ascribed to a range of other factors, including the use of different test passages and voices in Experiment 2. We return to this point in the General Discussion.

Although the findings of Experiments 1 and 2 are consistent, a potential confound that we must consider is that infants' responses in Experiment 2 could be based on gender matching (see, e.g., Poulin-Dubois, Serbin, & Derbyshire, 1998; Poulin-Dubois, Serbin, Kenyon, & Derbyshire, 1994; Richoz et al., 2017; Walker-Andrews et al., 1991). Did infants simply look at a female face when hearing a female voice and look at a male face when hearing a male voice? Our results from Experiment 1 do not support this idea. Here, this potential confound did not exist because all cartoons had gender-neutral appearances, and infants nonetheless recognized the talkers in the mixed-gender condition. However, to more confidently rule out potential effects of gender matching in the second experiment, we tested a new group of infants only on the test (but not the training) trials from the mixed-gender condition of Experiment 2. If our findings from Experiment 2 were driven by gender matching, then infants in Experiment 3 should match the faces and voices even without prior exposure to the talkers. However, if our findings were based on voice recognition, then infants should fail to match the faces and voices in Experiment 3.

EXPERIMENT 3

Experiment 3 explored the possibility that infants' high performance for the male/female talker pairs in Experiment 2 was based on audiovisual matching of gender cues rather than talker recognition per se. In Experiment 3, we presented a control group of 16.5-month-olds with the test (but not the training) trials from the mixed-gender condition of Experiment 2 and examined whether infants would look toward the gender-matched face when hearing a voice. Here, infants' looking behavior could not be guided by voice recognition because infants were not familiarized with the faces and voices prior to testing.

Method

Participants

Sixteen English-learning 16- to 17-month-olds (M_{age} = 502 days, range = 482–519; 10 female) from the Greater Toronto Area were tested (same eligibility criteria as in Experiments 1 and 2). Data for two additional infants were excluded due to fussing.

Stimuli

Stimuli consisted of the test trials from Experiment 2.

Procedure

The procedure was the same as in Experiment 2, except that infants in Experiment 3 were only tested on the test trials from the mixed-gender condition of Experiment 2. That is, instead of being familiarized with the talkers in a preceding training phase, infants were only presented with the test trials. We again measured infants' looking time to the face that matched the voice.

Coding and analysis

Data coding and analysis was the same as in Experiment 2.

Results and discussion

The data from Experiment 3 were analyzed using the same 2.5 sec window of analysis as reported for Experiments 1 and 2. If infants could match the faces and voices based on gender, then their proportion of looks to target should be significantly above chance (.5).

To establish whether our results in Experiment 2 were driven by gender detection, we analyzed the combined looking time data from Experiment 2 (mixed-gender condition) and Experiment 3 using a one-way ANOVA with training (training, no training) as independent variable. Results revealed a significant main effect of training on the proportion of looks to target, $F(1, 30) = 20.65$, $p < .001$, $\eta_p^2 = .41$, indicating that infants performed differently depending on whether or not they were exposed to the talkers before the test phase (see Figure 2, right panel, two rightmost bars). In contrast to Experiment 2, the proportion of looks to target did not significantly differ from chance in Experiment 3 ($M = .49$, $SD = .14$), $t(15) = -0.26$, $p = .795$.

We found no evidence that 16.5-month-old infants can map novel voices to gender-appropriate faces, which contrasts with earlier findings on gender matching in infancy (e.g., Poulin-Dubois et al., 1994, 1998; Richoz et al., 2017; Walker-Andrews et al., 1991). This may be ascribed to methodological differences across studies (regarding, e.g., trial length, use of dynamic or static faces, or prototypicality of the male and female faces and voices) and to the possibility that fewer visual gender cues were available to infants because the book covered the lower part of each talker's face. Critically, however, the results of our control experiment allow us to feel more confident that infants' performance in Experiment 2 was indeed driven by voice recognition and not simply by gender matching.

GENERAL DISCUSSION

Studying infants' ability to identify people by voice helps us to better understand how infants become adult-like in their processing of indexical information in speech, and how they process talker variability in language acquisition more generally. The developmental literature on talker processing conveys a somewhat contradictory message, with infants under 1 year of age performing well in talker processing tasks, but children in their school-age years struggling to identify talkers. One might even speculate that the development of children's talker recognition abilities follows a U-shaped curve (see also Creel & Quam, 2015). However, we found no evidence to support this idea. In three talker recognition experiments, we show that 16.5-month-old infants, much like school-age children, have difficulty learning to recognize unfamiliar same-gender talkers, suggesting that infants' talker recognition abilities are not superior to those of older children.

Our finding that 16.5-month-olds do not seem to have more sophisticated talker recognition skills than do older children allows us to reconcile the infant and early childhood literatures on talker processing. Previous infant studies have tested infants' ability to tell voices apart (e.g., Fecher & Johnson, 2018b; Friendly et al., 2014), and this type of task assesses short-term voice discrimination and can be performed without necessarily accessing higher-level linguistic information. In comparison, studies testing older children have mostly tested children's talker recognition skills (except, e.g., Levi, 2018; Levi & Schwartz, 2013), and talker recognition tasks involve later (encoding, storage, or retrieval) stages in perception and tax auditory memory capacity to a greater degree. These task differences across literatures not only complicate a comparison of infants' and children's talker processing skills, they are also suggestive of more advanced talker processing in infants than older children—a notion that was, however, not supported by the current data. Interestingly, similar task-based discrepancies in the infant and early childhood literatures have been reported for other areas of spoken language processing, such as speech sound acquisition, word learning, and processing of vocal affect (see Creel & Quam, 2015). The issues raised in the present paper are therefore representative of the age and task divides that appear to exist in developmental science more widely. As previously discussed in Creel and Quam (2015), an important next step is to bridge these divides by developing paradigms that allow us to link infants' early perceptual sensitivities (like talker discrimination) with children's (and adults') perceptual memory and associative learning abilities (like talker recognition). Our study is a first step in this direction. Thus, aside from advancing our understanding of talker recognition, we hope to spark similar debate in other domains of developmental investigation, even beyond the speech and language domain.

What does the finding that infants do not outperform children at talker recognition teach us about the developmental trajectory of talker recognition? It was previously suggested that the development of talker recognition might be discontinuous (Creel & Jiménez, 2012). But until now, developmental research has largely focused on talker recognition in later childhood. For example, Creel and Jiménez (2012) found that 3- to 6-year-old children are significantly worse at talker learning than adults, and from this inferred that children show protracted perceptual attuning to (rather than filtering out of) talker-related information in speech. Our data extend these findings by providing a first glimpse into unfamiliar talker recognition between infancy and childhood (to our knowledge, no other studies have examined talker recognition in 1- to 3-year-olds). Our findings provide no evidence of a discontinuity in development and therefore align with the protracted tuning hypothesis favored by Creel and Jiménez (2012). Moreover, our data support the idea that talker recognition requires more protracted learning than talker discrimination. That is, infants can successfully tell apart voices from a very young age, an ability which is refined throughout childhood (Levi, 2018; Levi & Schwartz, 2013; Mann et al., 1979) and which can be conceptualized as an important foundation for talker recognition. However, infants require years of perceptual learning before they reach mature levels of remembering voices or associating voices with specific individuals.

Establishing the developmental time course of talker recognition will not only enhance our understanding of indexical processing in infancy and early childhood, it will also inform our attempts to understand why it takes children so long to reach mature talker recognition abilities. One explanation for the finding that talker recognition undergoes protracted tuning across development is that children need to learn to tune in to the (supra)segmental speech cues that most effectively signal talker identity (when the goal is to determine who is speaking), and at the same time they need to learn to tune out irrelevant talker variation in speech (when the goal is to understand the spoken message). This perceptual learning process is complicated in at least two ways. First, the same speech cues may be important for determining both talker identity and speech content (e.g., the spectral composition of a fricative may signal a difference between two talkers as well as the distinction between the words "sip" and "ship"; e.g., Remez, Fellowes, & Nagel, 2007). And second, knowledge of language-specific speech sounds affects talker processing early on, as demonstrated by the language familiarity effect (e.g., Fecher & Johnson, 2018b; Johnson et al., 2011). Hence, studying when and how children develop the ability to recognize people by voice informs developmental models of talker recognition, and it helps us to better understand the relationship between talker recognition and spoken language processing more generally (see, e.g., Andics, McQueen, & Van Turennout, 2007; Goggin, Thompson, Strube, & Simental, 1991; Houston & Jusczyk, 2000; Johnson, Bruggeman, & Cutler, 2017; Levi & Schwartz, 2013; Mullennix, Pisoni, & Martin, 1989).

Finally, our data hint at the possibility that talker learning is facilitated when infants have access to visual speech cues. When we increased the naturalness and social relevance of our task by presenting infants with talking human faces instead of animated cartoon animals, infants showed a tendency to perform better (at least for the mixed-gender voices). We cannot draw any firm conclusions from our data because various factors may have contributed to this effect, including differences in the test passages and voices used in Experiments 1 and 2. However, we speculate that showing infants real people during voice learning was a pivotal factor in infants' increased talker recognition performance in Experiment 2. That is, perhaps voice learning was more successful in Experiment 2 because infants had access to visual speech cues that were not available from the cartoons. This idea is supported by research on audiovisual speech perception in infancy, which has shown that talking human faces are a particularly rich source of information about a person's identity given they provide both modality-specific information (e.g., facial appearance, voice pitch) and amodal information (e.g., rhythmic synchrony in the speech signal and the movements of the lips and jaw). Infants not only show preferences for faces and face-like visual configurations over other types of visual stimuli (e.g., Goren, Sarty, & Wu, 1975; Mondloch et al., 1999), they are also highly adept at integrating auditorily and visually presented speech. For example, infants readily map lip shapes onto vowel sounds (Patterson & Werker, 2003), are susceptible to the McGurk effect (Burnham & Dodd, 2004), match gender and emotional affect across faces and voices (Walker-Andrews, 1997; Walker-Andrews et al., 1991), and dishabituate to mismatched face-voice pairings (Bahrick et al., 2005; Brookes et al., 2001). Here, we tentatively conclude that talker recognition may improve when visual speech cues are available during encoding of the voices, and we encourage future research to corroborate this conclusion.

To conclude, this study examined infants' ability to identify people by voice. We found that unfamiliar same-gender talker recognition poses a complex cognitive challenge for 16.5-month-old infants. This suggests that unfamiliar talker recognition is mastered later in life than talker discrimination, and that task differences in the studies that have previously tested infants (on talker discrimination) and children (on talker recognition) account for the apparent talker recognition paradox in developmental speech perception. Testing a wider range of ages using a greater variety of tasks and stimuli will provide a richer account of talker learning in infancy and across child development. Future research will also be needed to address how infants learn to increase their sensitivity to indexical information in speech when the goal is to determine who is talking, and at the same time learn to ignore talker variation when it is detrimental to comprehending the spoken message. Research on these and related topics will advance our understanding of talker recognition and spoken language processing within and across modalities, and across the life span.

ACKNOWLEDGMENTS

Thanks to Lisa Hotson, Katrina Aranas, and Shukri Nur for their assistance with stimulus creation and data collection, and all participating families. This research was supported by grants awarded to E.K.J. from the Social Sciences and Humanities Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada, and the Canada Research Chairs Program. The authors declare no conflicts of interest with regard to the funding source for this study. Portions of this work were presented at the 20th International Conference on Infant Studies, New Orleans, Louisiana, USA (May 2016).

REFERENCES

Andics, A., McQueen, J. M., & Van Turennout, M. (2007). Phonetic context influences voice discriminability. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1829–1832). Saarbrücken, Germany.

Bahrick, L. E., Hernandez-Reif, M., & Flom, R. (2005). The development of infant learning about specific face-voice relations. *Developmental Psychology*, *41*, 541–552.

Beauchemin, M., González-Frankenberger, B., Tremblay, J., Vannasing, P., Martínez-Montes, E., Belin, P., ... Lassonde, M. (2011). Mother and stranger: An electrophysiological study of voice processing in newborns. *Cerebral Cortex*, *21*, 1705–1711.

Bricker, P. D., & Pruzansky, S. (1966). Effects of stimulus content and duration on talker identification. *The Journal of the Acoustical Society of America*, *40*, 1441–1449.

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* (1–2), 75–82.

Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk Effect. *Developmental Psychobiology*, *45*(4), 204–220.

Cook, S., & Wilding, J. (1997). Earwitness testimony: Never mind the variety, hear the length. *Applied Cognitive Psychology*, *11*, 95–111.

Creel, S. C., & Jiménez, S. R. (2012). Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology*, *113*, 487–509.

Creel, S. C., & Quam, C. (2015). Apples and oranges: Developmental discontinuities in spoken-language processing? *Trends in Cognitive Sciences*, *19*, 713–716.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*, 1174–1176.

DeCasper, A. J., & Prescott, P. A. (1984). Human newborns' perception of male voices: Preference, discrimination, and reinforcing value. *Developmental Psychobiology*, *17*, 481–491.

Fecher, N., & Johnson, E. K. (2018a). Effects of language experience and task demands on talker recognition by children and adults. *The Journal of the Acoustical Society of America*, *143*, 2409–2418.

Fecher, N., & Johnson, E. K. (2018b). The native-language benefit for talker identification is robust in 7.5-month-old infants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 1911–1920.

Floccia, C., Nazzi, T., & Bertoncini, J. (2000). Unfamiliar voice discrimination for short stimuli in newborns. *Developmental Science*, *3*, 333–343.

Friendly, R. H., Rendall, D., & Trainor, L. J. (2014). Learning to differentiate individuals by their voices: Infants' individuation of native- and foreign-species voices. *Developmental Psychobiology*, *56*(2), 228–237.

Goggin, J. P., Thompson, C. P., Strube, G., & Simental, L. R. (1991). The role of language familiarity in voice identification. *Memory and Cognition*, *19*, 448–458.

Goren, C. C., Sarty, M., & Wu, P. V. K. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, *56*, 544–549.

Hepper, P. G., Scott, D., & Shahidullah, S. (1993). Newborn and fetal response to maternal voice. *Journal of Reproductive and Infant Psychology*, *11*(3), 147–153.

Hollich, G. (2005). *Supercoder: A program for coding preferential looking (Version 1.5)*. West Lafayette, IN: Purdue University.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1570–1582.

Johnson, E. K., Bruggeman, L., & Cutler, A. (2017). Abstraction and the (misnamed) language familiarity effect. *Cognitive Science*, *42*(2), 633–645.

Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, *14*, 1002–1011.

Kisilevsky, B. S., & Hains, S. M. J. (2011). Onset and maturation of fetal heart rate response to the mother's voice over late gestation. *Developmental Science*, *14*(2), 214–223.

Kisilevsky, B. S., Hains, S. M. J., Brown, C. A., Lee, C. T., Cowperthwaite, B., Stutzman, S. S., … Wang, Z. (2009). Fetal sensitivity to properties of maternal speech and language. *Infant Behavior and Development*, *32*(1), 59–71.

Langlois, J. H., Roggman, L. A., Casey, R. J., Ritter, J. M., & Rieser-Danner, L. A. (1987). Infant preferences for attractive faces: Rudiments of a stereotype? *Developmental Psychology*, *23*, 363–369.

Lecanuet, J.-P., Granier-Deferre, C., Jacquet, A.-Y., Capponi, I., & Ledru, L. (1993). Prenatal discrimination of a male and a female voice uttering the same sentence. *Early Development and Parenting*, *2* (4), 217–228.

Levi, S. V. (2018). Another bilingual advantage? Perception of talker-voice information. *Bilingualism: Language and Cognition*, *21*, 523–536.

Levi, S. V., & Schwartz, R. G. (2013). The development of language-specific and language-independent talker processing. *Journal of Speech, Language, and Hearing Research*, *56*, 913–920.

Mann, V. A., Diamond, R., & Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology*, *27*(1), 153–165.

Mehler, J., Bertoncini, J., Barrière, M., & Jassik-Gerschenfeld, D. (1978). Infant recognition of mother's voice. *Perception*, *7*, 491–497.

Miller, C. L. (1983). Developmental changes in male/female voice classification by infants. *Infant Behavior and Development*, *6*(2–3), 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*(2–4), 143–159.

Mills, M., & Melhuish, E. (1974). Recognition of mother's voice in early infancy. *Nature*, *252*, 123–124.

Mondloch, C. J., Lewis, T. L., Budreau, D. R., Maurer, D., Dannemiller, J. L., Stephens, B. R., & Kleiner-Gathercoal, K. A. (1999). Face perception during early infancy. *Psychological Science*, *10*, 419–422.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America*, *85*(1), 365–378.

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological Science*, *15* (2), 133–137.

Öhman, L., Eriksson, A., & Granhag, P. A. (2011). Overhearing the planning of a crime: Do adults outperform children as earwitnesses? *Journal of Police and Criminal Psychology*, *26*(2), 118–127.

Paquette-Smith, M., & Johnson, E. K. (2016). I don't like the tone of your voice: Infants use vocal affect to socially evaluate others. *Infancy*, *21*(1), 104–121.

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, *6*(2), 191–196.

Perea, M., Jiménez, M., Suárez-Coalla, P., Fernández, N., Viña, C., & Cuetos, F. (2014). Ability for voice recognition is a marker for dyslexia in children. *Experimental Psychology*, *61*, 480–487.

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.

Purhonen, M., Kilpeläinen-Lees, R., Valkonen-Korhonen, M., Karhu, J., & Lehtonen, J. (2004). Cerebral processing of mother's voice compared to unfamiliar voice in 4-month-old infants. *International Journal of Psychophysiology*, *52*, 257–266.

Remez, R. E., Fellowes, J. M., & Nagel, D. S. (2007). On the perception of similarity among talkers. *The Journal of the Acoustical Society of America*, *122*, 3688–3696.

Richoz, A.-R., Quinn, P. C., Hillairet de Boisferon, A., Berger, C., Loevenbruck, H., Lewkowicz, D. J., … Pascalis, O. (2017). Audio-visual perception of gender by infants emerges earlier for adult-directed speech. *PLoS One*, *12*(1), e0169325.

Sadakata, M., & McQueen, J. M. (2013). High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates. *The Journal of the Acoustical Society of America*, *134*(2), 1324–1335.

Stevenage, S. V. (2017). Drawing a distinction between familiar and unfamiliar voice processing: A review of neuropsychological, clinical and empirical findings. *Neuropsychologia*, *116*(Pt B), 162–178.

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.

Turnure, C. (1971). Response to voice of mother and stranger by babies in the first year. *Developmental Psychology*, *4*(2), 182–190.

Van Lancker, D., & Kreiman, J. (1987). Familiar voice recognition and unfamiliar voice discrimination are independent and unordered abilities. *Neuropsychologia*, *25*, 829–834.

Voegtline, K. M., Costigan, K. A., Pater, H. A., & DiPietro, J. A. (2013). Near-term fetal response to maternal spoken voice. *Infant Behavior and Development*, *36*, 526–533.

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin*, *121*, 437–456.

Walker-Andrews, A. S., Bahrick, L. E., Raglioni, S. S., & Diaz, I. (1991). Infants' bimodal perception of gender. *Ecological Psychology*, *3*(2), 55–75.

Ward, C. D., & Cooper, R. P. (1999). A lack of evidence in 4-month-old human infants for paternal voice preference. *Developmental Psychobiology*, *35*(1), 49–59.

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, *3*(1), 1–30.

Winters, S. J., Levi, S. V., & Pisoni, D. B. (2008). Identification and discrimination of bilingual talkers across languages. *The Journal of the Acoustical Society of America*, *123*, 4524–4538.

Yarmey, A. D., & Matthys, E. (1992). Voice identification of an abductor. *Applied Cognitive Psychology*, *6*, 367–377.

APPENDIX A

TRAINING AND TEST PASSAGES

Training passages in Experiments 1–3 and test passages in Experiment 1

1 You need to put on your jacket and mittens, it's cold outside.
2 Let me tie your shoe. You're going to trip and fall with it untied.
3 Go over and thank the nice lady for helping you find your coat.
4 You have cookie crumbs all over your face, let me wash them off.
5 We read this book last week, but we can read it again if you want to.
6 Put all your toys away, we need to tidy up before dad gets home.
7 It's ok, we can come back and play at the park tomorrow.
8 If you sit really quietly now, then we can get ice cream after.
9 You have to put back the rubber duck. We have one just like it at home.
10 Don't put that in your mouth, you don't know where it has been.

Test passages in Experiments 2 and 3

1 Can you find me? Over here! Look at me!
2 Look who's hiding! Oh, it's me! Look at me!
3 Where am I? Oh, here I am! Look at me!
4 I'm here behind the book! See? Look at me!

APPENDIX B

ADDITIONAL TALKER CHARACTERISTICS

TABLE A1 Acoustic-Phonetic Measurements (Fundamental Frequency [F0], Articulation Rate) for All Male (M) and Female (F) Talkers in the Mixed-Gender and Same-Gender Conditions of Experiments 1–3 (Child-Directed Speech)

Experiment	*Condition*	*Talker pair*	*Talker*	*Age (in years)*	*F0 M and SD (in Hertz)*	*Articulation rate (in syllables/sec)*
1	Mixed gender	1	M1	21	200.5 (*45.8*)	3.7
			F1	20	271.6 (*53.7*)	3.7
		2	M2	19	194.5 (*45.8*)	3.7
			F2	20	277.1 (*61.7*)	4.4
	Same gender	1	F3	21	257.1 (*63.4*)	4.3
			F1	20	271.6 (*53.7*)	3.7
		2	F4	37	249.9 (*65.9*)	3.9
			F5	20	247.8 (*58.5*)	4.3
2 and 3	Mixed gender	1	M3	23	165.8 (*46.6*)	4.3
			F6	24	247.1 (*62.7*)	3.6
		2	M4	25	149.6 (*41.3*)	4.0
			F7	21	277.1 (*65.2*)	4.3
	Same gender	1	F8	31	200.6 (*45.5*)	4.1
			F6	24	247.1 (*62.7*)	3.6
		2	F9	22	253.2 (*44.1*)	3.9
			F7	21	277.1 (*65.2*)	4.3

Notes. The table shows each talker's age at time of recording and how the talkers were paired across experiments. It was ensured that the talkers within a pair were well matched in vocal affect (which led us to replace F2 with F5 in Experiment 1). All acoustic measurements are based on child-directed speech.