COLLOQUIUM PAPER

Two are better than one: Infant language learning from video improves in the presence of peers

Sarah Roseberry Lytle[a,1], Adrian Garcia-Sierra[b], and Patricia K. Kuhl[a]

[a]Institute for Learning & Brain Sciences, University of Washington, Seattle, WA 98195; and [b]Speech, Language, and Hearing Sciences, University of Connecticut, Storrs, CT 06269

Edited by David E. Meyer, University of Michigan, Ann Arbor, MI, and approved November 29, 2017 (received for review May 11, 2017)

Studies show that young children learn new phonemes and words from humans significantly better than from machines. However, it is not clear why learning from video is ineffective or what might be done to improve learning from a screen. The present study, conducted with 9-month-old infants, utilized a manipulation—touch screen video—which allowed infants to control presentations of foreign-language video clips. We tested the hypothesis that infant learning from a screen would be enhanced in the presence of a peer, as opposed to learning alone. Brain measures of phonetic learning and detailed analyses of interaction during learning confirm the hypothesis that social partners enhance learning, even from screens.

language learning | phoenemic discrimination | screen media | social learning | infancy

Many studies have shown that young children's learning of language material from video is very low compared with learning from human tutors, a pattern called the "video deficit." (1) For example, research has established infants' ability to learn foreign language phonemes (the consonants and vowels that make up words) through social but not nonsocial contexts (2). In Kuhl et al. (2), 9-mo-old infants were exposed to Mandarin Chinese in 12 25-min laboratory visits. Each infant experienced one of three exposure styles: live social presentation, the same foreign speakers and material presented on video, or an audio recording of the same speakers and material. A control group of infants experienced live language social presentation but heard only English. Phonemic learning, tested with behavioral and brain measures after completion of the second language (L2) exposure sessions, demonstrated that only infants exposed to live Mandarin speakers discriminated the foreign phonemes as well as native Mandarin-learning infants; no learning occurred when exposure occurred through video displays or audio recordings (2).

Other studies confirm that children's language learning is better from live humans than from screens (3–5). In one study, video clips from *Sesame Beginnings* presented novel verbs to 2.5- and 3-y-old children (5). Half of the children saw the novel verbs presented entirely on video; the other half saw a 50–50 split of presentations on video and delivered by a live social partner. Children were tested on their ability to extend the novel verb to a new actor performing the same action. Results showed that toddlers who interacted with an adult in addition to watching a video learned the novel verbs at a younger age than children who passively viewed the video, and that learning from video was not as robust as learning from live social interactions.

Recent evidence suggests that the screen itself does not impede children's learning; rather, the problem is the lack of interactivity in traditional media. One study used video chats to ask if 24- to 30-mo-olds can learn language in a video context that incorporates social interactions (6). Even though video chats offer a 2D screen, this technology differs from traditional video in several important ways. Video chats allow children and an adult to participate in a two-way exchange, thereby approximating live social interactions. Adults are also able to be responsive to children and ask questions that are relevant to them. Although the speaker's eye gaze is often distorted in video chats because of the placement of the camera relative to the screen, video chats preserve many of the qualities of social interactivity that help children learn (7). In fact, when 24- to 30-mo-olds were exposed to novel verbs via video chat, children learned the new words just as well as from live social interactions. Toddlers showed no evidence of learning from noninteractive video. Myers et al. (8) recently demonstrated a similar phenomenon with 17- to 25-mo-olds. These young toddlers experienced either a FaceTime conversation or a prerecorded video of the same speaker. A week after exposure, children who interacted with an unknown adult via FaceTime recognized the adult and demonstrated word and pattern learning. Thus, research provides evidence that children's ability to learn language from screens can be improved by technology that facilitates social interactions (e.g., video chats) (6, 8), by the content of media (e.g., reciprocal social interactions) (9), or with the context of screen media use (e.g., coviewing) (10). This allows the field to move beyond the screen vs. live dichotomy and focus the discussion on the role of interactivity for children's learning.

Historically, research on the effect of social interactivity in children's media has always paired adults with children. However, there is some evidence that infants may also treat peers as social partners. For example, Hanna and Meltzoff (11) found that 14- and 18-mo-olds imitate actions demonstrated by same-aged peer models and even recall these actions after a delay of 2 d. More recent evidence even suggests a peer advantage, such that 14- and 18-mo-olds imitated complex action sequences better from 3-y-old models than from adult models (12).

Additional research indicates that children's learning is enhanced in the mere presence of others. That is, even when a peer is not serving as a teacher or model, simply being in the presence of a social other may facilitate learning. Studies show that infants perform tasks differently—and better—when they are in the presence of another person (13), and research with school-aged children suggests that learning is improved by the mere presence of another person, or even the illusion that another person is present (14). These findings indicate that explorations of the effects of having a social partner on children's learning from media might broaden our understanding of the roles played by social peers, even if peers are not in the position of a teacher or model.

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Digital Media and Developing Minds," held October 14–16, 2015, at the Arnold and Mabel Beckman Center of the National Academies of Sciences and Engineering in Irvine, CA. The complete program and video recordings of most presentations are available on the NAS website at www.nasonline.org/Digital_Media_and_Developing_Minds.

Author contributions: S.R.L. and P.K.K. designed research; S.R.L. and A.G.-S. performed research; S.R.L. and A.G.-S. analyzed data; and S.R.L., A.G.-S., and P.K.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: All relevant data are stored on a server at the Institute for Learning and Brain Sciences at the University of Washington and are available at http://ilabs.uw.edu/interactive-exposure-2017.

[1]To whom correspondence should be addressed. Email: sarahr28@uw.edu.

Published online October 1, 2018.

The current study investigates the effect of the presence of peers on infant foreign-language phonetic learning from video. We utilized the same Mandarin-language videos used previously in passive learning experiments (2), but made the procedure an active learning environment by allowing infants to control the presentation of videos using a touch screen. Each touch of the screen initiated a 20-s clip of the Mandarin speaker talking about toys and books. Given that previous research on children's learning from peers or in the presence of peers presents children with a task (11–14), the touchscreen paradigm gives infants an active learning task that nevertheless presents the same information as previous research (2). We manipulate the presence of peers by randomly assigning infants to an individual-learning condition or a paired-learning condition. Infants in the individual condition participated in all study sessions by themselves, whereas infants in the paired condition always participated with another infant (Fig. 1).

To measure infant's foreign-language sound discrimination, we employ a behavioral measure, "conditioned head turn," as well as a brain measure, event related potentials (ERPs). Kuhl et al. (2) reported results based on a conditioned head turn paradigm, but ERPs are also commonly used to assess infants' ability to discriminate the sounds of language (15–17). Specifically, previous ERP investigations have shown that adults (18, 19) as well as infants (20–23) exhibit the characteristic mismatch negativity (MMN), a negative-polarity waveform that occurs about 250–350 ms after the presentation of the deviant sound, indicating neural discrimination of the change from one phonetic unit to the other. Importantly, the MMN is elicited in adults and 10- to 12-mo-old infants when listening to sounds of their native language, and it is reduced or absent when they listen to speech sounds that do not represent phonemic categories in their native language (24–26). Thus, we employ both measures of speech discrimination in the present study.

We hypothesize that the mere presence of peers in the present investigation will support children's ability to discriminate the foreign-language sounds. During the exposure visits, the effect of peers will be evident in children's social behavior, and we expect that social cues, like vocalizations and eye gaze, indicators of early attention (27, 28) and communication skills (29–31), will emerge as related to infant phonemic learning. With regard to the measures of sound discrimination, behavioral evidence of discrimination will be evidenced by performance greater than chance in the conditioned head turn paradigm and through the presence of the MMN in the ERP test. Infant research has shown that attention plays a role in the generation of the MMN. Specifically, auditory change detection can occur with high or low attentional demands that are mediated by language experience (15–17, 25, 32–34), discriminability of the signals (35–37), and maturational factors (38–42). The MMN associated with high attentional demands in the perception of speech sounds exhibits a positive polarity (positive-MMR or pMMR) and is considered a less-mature MMN response. The MMN associated with low attentional demands exhibits a negative polarity (i.e., MMN) and, because it is shown by adults listening to native-language sounds, is considered the more mature MMN. In the present investigation, we postulate that infants in the single-infant condition will show pMMRs because high attentional demands are required by a difficult speech discrimination task, such as nonnative speech discrimination (35, 36, 43). On the other hand, we postulate that infants in the paired-infant condition will process the Chinese phonetic distinction with less effort due to social arousal, and hence we expect the brain response to have a negative polarity (i.e., MMN).

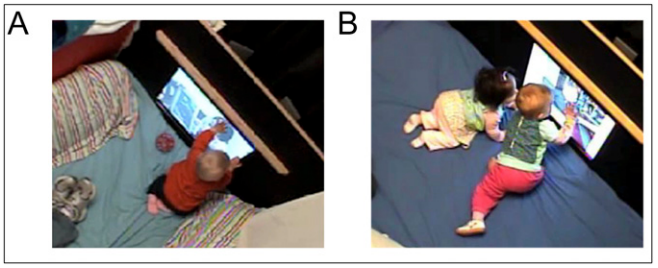

Fig. 1. Examples of the individual- (*A*) and paired- (*B*) exposure sessions.

Results

Preliminary analyses were conducted to examine basic elements of the exposure sessions. Because the procedure allowed for flexibility in the length of sessions, and sessions were terminated based on criteria that were not time-based, we first examined the length of the individual- and paired-exposure sessions. Individual and paired sessions were equally long (individual sessions mean = 20.61 min, SD = 8.32; paired sessions mean = 17.26 min, SD = 1.85), $t(29) = 1.53$, $P > 0.05$. Additionally, infants viewed the same number of videos across conditions (individual sessions mean = 26.69 videos, SD = 9.64; paired sessions mean = 23.37 videos, SD = 3.56), $t(29) = 1.26$, $P > 0.05$. Finally, the rate at which infants watched videos in the individual-exposure sessions and in the paired-exposure session did not differ (individual sessions mean = 1.35 videos per minute, SD = 0.46; paired sessions mean = 1.37 videos per minute, SD = 0.27), $t(29) = 0.09$, $P > 0.05$.

We next examined the results from the two tests of phonemic learning, the conditioned head-turn task and the ERP test of discrimination. Whereas the head-turn task provided a behavioral test of discrimination, the ERP test provided a neural test of discrimination. Infants in the present study did not show behavioral evidence of learning foreign language phonemes regardless of exposure condition (individual sessions mean = 50.66%, SD = 6.69; paired sessions mean = 50.61%, SD = 9.39). These results for individual and paired sessions did not differ from the chance (50%), $t(12) = 0.36$, $P > 0.05$ and $t(13) = 0.24$, $P > 0.05$, respectively, nor did they differ from each other, $t(25) = 0.02$, $P > 0.05$.

Next, the ERP data were submitted for analysis. Infants' ERP mean amplitudes to the standard and deviant sound were analyzed in two time windows (150–250 ms and 250–350 ms) (Fig. 2). For the 150- to 250-ms time window, infants in the paired group showed no significant difference between standard and deviant ERP responses at either electrode site: Fz, $F(1, 14) = -0.93$, $P = 0.364$; Cz, $F(1, 14) = -1.62$, $P = 0.127$. In contrast, infants in the individual group showed significant differences at both electrode sites: Fz, $F(1, 15) = -2.56$, $P = 0.022$; Cz, $F(1, 15) = -2.41$, $P = 0.029$. For the 250- to 350-ms time window, infants in the paired group showed no significant difference between standard and deviant ERP responses for Fz, $F(1, 14) = 1.88$, $P = 0.081$, but a significant difference was found at Cz electrode, $F(1, 14) = 3.14$, $P = 0.007$. Infants in the individual group showed no significant differences at either electrode site for the late time window: Fz, $F(1, 15) = -1.56$, $P = 0.14$; Cz, $F(1, 15) = -1.14$, $P = 0.27$.

Fig. 2 shows a waveform with negative polarity, an MMN, in response to the deviant in the 250- to 350-time window for the paired group, which peaked at 300 ms (see central electrode sites Fz and Cz in Fig. 2). Infants in the individual group did not exhibit this negativity, and instead exhibited a positive polarity waveform, a pMMR, in response to the deviant in the earlier 150- to 250-ms time window (see central electrode sites Fz and Cz in Fig. 2). Differences between individual and paired groups were significant at both electrode sites: Fz, $F(1, 29) = 6.059$, $P = 0.02$; Cz, $F(1, 29) = 10.754$, $P = 0.003$. There was no correlation

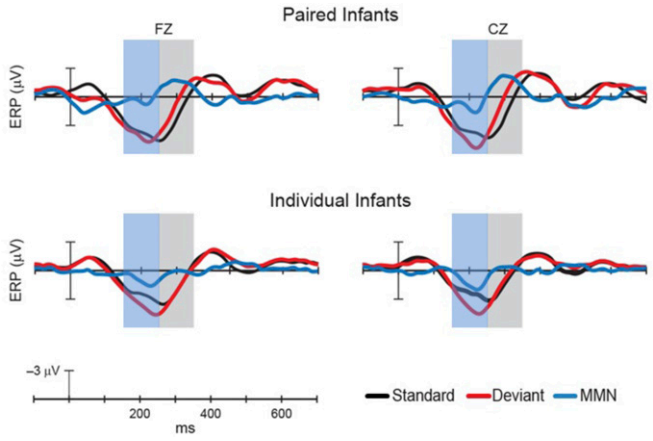

Fig. 2. Differences between individual and paired groups were significant at both electrode sites: Fz, $F(1, 29) = 6.059$, $P = 0.02$, and Cz: $F(1, 29) = 10.754$, $P = 0.003$. Note that the ERP is plotted with negative voltage upward.

between infants' performance on the conditioned head-turn task and their ERP data (rs < 0.2, Ps > 0.33).

To further explore the differences between exposure conditions, and to determine the causal mechanism of paired exposure success, we examined several measures of infant behavior during exposure sessions. First, we examined infant touches to the touchscreen. Although the number of overall touches to the screen, including those that triggered videos and those that occurred while videos were playing, did not differ across condition [individual-exposure sessions mean = 224.83 touches per session, SD = 108.42; paired-exposure sessions mean = 231.45 touches per session, SD = 56.84; $t(29) = 0.21$, $P > 0.05$], both infants in the paired condition contributed to the touch count. Therefore, infants in the individual exposure sessions touched the screen more than did each infant in the paired-exposure condition [individual-exposure sessions mean = 224.83 touches per session, SD = 108.42; paired-exposure sessions mean = 113.69 touches per session, SD = 54.83; $t(29) = 3.56$, $P = 0.001$]. Of the total number of touches, the vast majority of touches were intentional, not accidental, [individual-exposure sessions = 95.79% intentional touches, SD = 0.08; paired-exposure sessions = 97.37%, SD = 0.03; $t(29) = 0.72$, $P > 0.05$]. Finally, infants in the two conditions did not differ in their motor skills at the beginning, $t(29) = 0.10$, $P > 0.05$, the middle, $t(29) = 0.35$, $P > 0.05$, or at the end of the exposure visits, $t(29) = 0.45$, $P = 0.65$. Infants touched the screen equally regardless of motor ability (rs < 0.04, Ps > 0.82). Coders often noted that caregivers of infants with less-developed motor skills would position the infants close to the touchscreen, eliminating the need for infants to have developed motor skills to touch the touchscreen.

Next, we compared infant vocalizations between children in each of the exposure conditions. Infants in the paired-exposure condition made marginally more speech-like vocalizations than infants in the individual-exposure condition, $t(29) = 1.89$, $P = 0.068$. Moreover, greater numbers of speech-like vocalizations were correlated with a negative MMN response in the 250- to 350-ms time interval, $r = -0.36$, $P < 0.05$. While paired infants produced greater numbers of speech-like vocalizations compared with babies in the individual group, nonspeech vocalizations did not differ between conditions, $t(29) = 1.42$, $P > 0.05$. Interestingly, infants who produced more speech-like vocalizations also scored higher on the conditioned head-turn task, $r = 0.43$, $P < 0.05$.

With regard to infant eye gaze during the exposure sessions, infants in the individual-exposure sessions looked at their own caregiver longer than infants in the paired-exposure condition (individual-exposure sessions mean = 58.20 s, SD = 21.82; paired-exposure sessions mean = 42.46 s, SD = 17.14), $t(29) = 2.22$, $P < 0.05$. Infants in individual sessions also looked to the screen longer than infants in the paired-exposure sessions (individual-exposure sessions mean = 124.66 s, SD = 33.32; paired-exposure sessions mean = 52.74 s, SD = 19.37), $t(29) = 7.28$, $P < 0.001$. Greater eye gaze toward the infant's own caregiver or the screen did not correlate with the MMN response in the 250- to 350-ms time interval, rs < 0.23, Ps > 0.23. For children in the paired-exposure sessions, we also coded looks to the other baby and to the other caregiver. Interestingly, children in the paired condition looked longest at the other baby (mean = 70.42 s, SD = 24.79), marginally longer than they looked toward the screen, $t(15) = 2.04$, $P = 0.059$. In contrast, children's looking time toward the other caregiver was the shortest (mean = 19.92 s, SD = 8.74), and significantly shorter than their looks toward their own caregiver, $t(15) = 3.93$, $P = 0.001$.

Finally, we examined learning as a function of the number of different learning partners that infants in the paired-exposure conditions experienced over the course of the L2 exposure sessions. Learning increased significantly as a function of the number of unique partners experienced by paired infants, both in relation to the MMN at electrode Cz ($r = -0.53$, $P < 0.05$) and in terms of a composite measure of MMN across Fz and Cz electrode sites ($r = -0.52$, $P < 0.05$) (Fig. 3). We also find that infants paired with more unique partners over the course of the exposure sessions produced more speech-like vocalizations, $r_s = 0.68$, $P < 0.05$.

Discussion

The present study examined the presence of peers on infants' early phonemic discrimination. We hypothesized that peers would support infants' ability to discriminate the foreign-language sounds, as indicated by behavioral and neural tests of speech discrimination, and that social cues, like vocalizations and eye gaze, would be related to infant phonemic discrimination. The results of the study revealed brain-based evidence of immature phonemic learning in infants in the individual-exposure sessions, while evidence of more mature learning emerged from infants in the paired-exposure sessions. Infants in the individual-exposure sessions exhibited a positive mismatch response, indicating increased cognitive demands during the auditory processing of the Chinese contrast (15, 44, 45). The negative polarity in the MNN

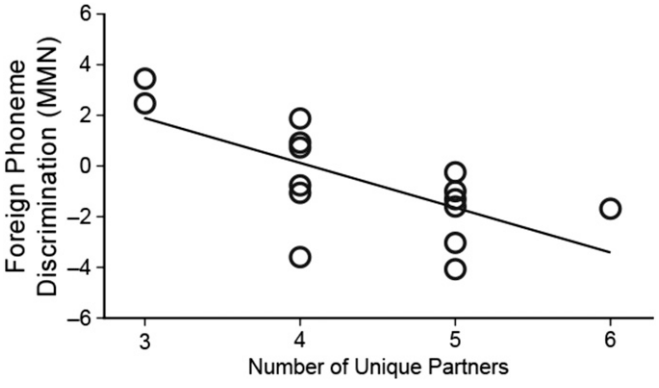

Fig. 3. Among infants in the paired baby condition, those who were partnered with more babies showed the strongest evidence of learning. In paired sessions, infants with more partners had a greater ERP mismatch negativity, $r_s = 0.68$, $P < 0.05$, which is evidence of greater foreign language phoneme learning.

response exhibited by infants in the paired-exposure sessions suggests lower cognitive demands in the processing of the speech contrast. Critically, these differences could not be attributed to the amount of exposure time, the number of videos viewed, touches to the touchscreen, or infants' motor ability. Differences between groups did emerge in that infants in the individual-exposure sessions looked more at their own caregiver and to the screen than infants in the paired-exposure sessions.

Infants in the paired-exposure sessions produced more speech-like vocalizations than infants in the individual-exposure sessions. Moreover, a greater number of speech-like vocalizations in paired group infants was correlated with more mature (adult-like) MMN brain responses and also correlated with behavioral performance on the speech discrimination task. Finally, with regard to infants in the paired-exposure sessions, infants who were paired with more unique partners over the course of the 12 exposure sessions vocalized more and demonstrated more mature brain processing of speech, suggesting more sophisticated and robust learning (15, 34).

We hypothesized that infants in the paired condition would demonstrate both behavioral and neural evidence of phonetic discrimination, yet we only found neural evidence of discrimination. The pattern of neural response exhibited by infants in the paired-exposure condition is indicative of more mature brain processing of the sounds, and it is likely that these findings represent the earliest stages of infants' sound learning. Such a pattern has precedent in the literature. As Tremblay et al. (46) suggest, "speech-sound learning occurs at a preattentive level, which can be measured neurophysiologically (in the absence of a behavioral response) to assess the efficacy of training." In contrast to previous research (2), in which infants demonstrated phonemic discrimination via behavioral tests, infant discrimination here may have been too fragile to lead to a behavioral response. This pattern may provide additional evidence for the video deficit, as behavioral evidence of phonemic discrimination resulted from a live tutoring condition, whereas neural—but not behavioral—evidence of discrimination in the present study was from contingent screens. Importantly, however, a preattentive response in a phonetic discrimination task has important implications for later language learning. Previous research has shown a relationship between phonetic discrimination, measured neurally in infancy, and children's language skills at 24 and 30 mo of age (47).

The present results support previous findings highlighting the importance of social interaction for children's learning, especially for learning from screen media. Although the importance of social interactions is clear, research exploring the underlying mechanisms has been sparse. Theoretical accounts suggest that social interactions may benefit children's learning for two reasons (48). (*i*) Motivation differs in social and nonsocial settings because social partners increase arousal; data suggest that even minimal social connections to another person increase motivation (49). (*ii*) Information also varies in social and nonsocial settings; eye gaze and other socially delivered cues, shown to assist infants' learning of the meanings of words (50), are less available on video. The two theoretical accounts are not mutually exclusive; both can be operating in natural learning settings.

In the present study, infants in both groups had access to some informational and some motivational cues. However, the paired condition was designed to provide additional motivation in the form of a peer partner to infants. We interpret the present results as support for the motivational mechanism of social interactivity.

In terms of the information that infants were able to recruit, infants in both exposure conditions had access to a variety of social cues that could have provided useful information for language learning. During all exposure visits, caregivers were in the room with infants and infants may have attempted to recruit informational cues from their caregiver's eye gaze. Although infants looked to caregivers, particularly their own caregiver, this category does not account for the majority of infant looks in either condition. In the paired-exposure conditions, infants also had access to informational cues from the other baby and the other parent. Indeed, infants in the paired-exposure conditions spent more time looking at the other baby than to any other coded region. It may be that the infants learned how to activate the touchscreen from their partner baby or benefitted from seeing how their partner baby behaved in this situation. This conclusion aligns with the imitation literature on learning from peers, although in those cases, the children were older and the peers were trained to model particular actions (11, 12). Even though the peer babies in the present study were not trained, it may be that they nevertheless provided their partner babies with information that led to enhanced learning.

Another potential source of information in the current situation was the screen. Children in the individual exposure condition, in particular, attended to the screen. Although research suggests that socially delivered cues like eye gaze are less available on screen (6), it is possible that infants recruited enough information on the screen to show evidence of immature phonemic discrimination.

In support of the motivational mechanism, infants in both exposure conditions were required to interact and engage with the touchscreen. This differs from traditional screen media viewing and from the video presentation used by Kuhl et al. (2) in that children were active, not passive participants in the learning process. Active engagement has been identified as one of the pillars from the science of learning that also supports children's learning from screens (51). Active engagement in the present study put infants in control of their video-viewing experience. Infants were quick to learn that they had to touch the screen to activate a video. This agency infused active engagement in the learning environment.

In addition to engaging children, the touchscreen also provided temporal contingency between infants' touches and video activation. Infants are drawn to contingent responses early in life. At 4 mo of age, for example, infants prefer adults who respond contingently to adults who do not (52). Similarly, 8-mo-olds produce more mature vocalizations in response to contingent behavior from adults, but do not refine their vocalizations after noncontingent responses (53). Including contingency in the present design may have increased infants' ability to discriminate the phonemes. Although there is some evidence that toddlers learn more from touchscreens when they are prompted to touch a particular location on the screen that is relevant to the learning objectives (54), other studies find that toddlers who press a button to advance a computer demonstration are more likely to learn than those who passively watch a video (55). The present study simply required infants to touch any location on the screen to advance the video, which is a developmentally appropriate design for infants. That infants in both conditions showed some evidence of phonetic discrimination suggests that the contingent and engaging touchscreen may have increased children's motivation to learn in the social environment.

The stronger argument in support of the motivational mechanism of social interactivity is the mature learning exhibited by children in the paired-exposure condition. Here, multiple factors indicate that motivation, or social arousal, may have driven infants' phonemic discrimination. First, having access to a peer might have triggered the mere presence effect on children's learning. The privileged role of peers has been documented with older children (13, 14), but it's possible that infants also perform better and learn more in the presence of similarly aged peers. Second, infants in the paired-exposure condition produced more speech-like vocalizations. Infants are known to vocalize more when they are aroused (56) and the present results link increased

vocalizations to higher scores on the behavioral measure of phonemic discrimination and to a more mature (greater negative polarity) neural response, the MMN, a measure of learning that is shown in adults when responding to native-language contrasts (18–23). Finally, we discovered a relationship between the number of partners a given infant was paired with during his or her exposure sessions and their speech-like vocalizations, as well as a relationship between the degree of learning infants in the paired condition evidenced later at test. Previous research has demonstrated that novelty heightens arousal, whereas familiarity dampens arousal (57). Thus, infants who experienced more novelty with their peer partners were more likely to maintain a higher level of arousal during the exposure sessions. Heightened social arousal may be the underlying mechanism of the motivational hypothesis that increased infant learning from screens in the present study.

In sum, the present investigation demonstrates enhanced infant discrimination of the phonemes of a foreign language from touchscreen video when learning in the presence of a peer as opposed to in isolation. Enhanced phonetic discrimination in pairs was not a function of the number of videos viewed by the infants, the amount of exposure to the foreign language, the number of screen touches, or infants' attention to the screen. Infants paired with other infants while learning not only show enhanced brain measures when reacting to the foreign-language phonemes, but also show significantly higher numbers of speech vocalizations, which itself suggests increased learning and arousal. We interpret the data as indicating that infants are socially aroused in the presence of a peer, that their arousal increases when paired with a novel infant, and that this leads to enhancement of early phonemic discrimination. The study breaks ground by demonstrating that social arousal may play a role in early language learning.

Methods

Participants. Written informed consent was provided for the human subjects protocol. The University of Washington approved the protocol. Thirty-one 9-mo-old infants (mean = 9.27 mo; SD = 0.17; range = 9.0–9.82 mo) completed the study. Infants were randomly assigned to individual (n = 16; mean = 9.25; SD = 0.22; range = 9.0–9.82) or paired (n = 15; mean = 9.29; SD = 0.10; range = 9.10–9.39) conditions. The numbers of females did not differ from chance in either condition [n = 6 in individual, n = 9 in paired groups; $\chi(1) = 1.57$, $P = 0.21$]. When possible, infants in the paired condition were partnered with other probands in the paired condition, although scheduling did not always allow for this arrangement. In cases when a partner infant from the experimental group was unavailable, infants were paired with substitute babies who were otherwise identical to the probands (n = 16). Data were not collected from these substitute babies. An additional 11 participants were excluded from the current dataset for an incomplete set of exposure visits (4 participants) and incomplete ERP test data (7 participants). All infants were full-term and were from monolingual English-speaking households.

Design. To investigate the mechanisms that drive language learning in social situations, the present study modified the design introduced by Kuhl et al. (2) that tested infants' learning of foreign language phonemes after a series of exposure visits. As before, infants visited the laboratory for a series of 12 exposure visits over 4 wk. During these sessions, infants were exposed via interactive touch screen video to 20-s clips of a Mandarin speaker talking about books and toys. Videos contained the target Mandarin phonemes, /tɕʰi/and (/ɕi/), a Mandarin Chinese consonant contrast that is not distinguished in English. After exposure, infants were assessed on phonemic learning in two ways. First, infants participated in a conditioned head turn task in which they learned to turn their head toward a loudspeaker when they detected a target phoneme interspersed among the background, or standard sounds (58). Second, infants' learning was also tested using ERPs (59). Together, these tasks provided both behavioral and neural measures of learning.

Apparatus. Exposure visits took place in a small room that was designed to be infant-friendly. A queen-sized memory foam mattress covered the floor and pillows were arranged along the sides of the room. At the front of the room, a 24.25-inch touchscreen was encased in a wooden façade that extended the width of the room and positioned the screen three inches higher than the floor. The wooden façade also had a small wooden shelf above the screen that was designed to allow infants to pull themselves up to a standing position in front of the screen. For paired-infant exposure sessions, the entire room was available to the infants. For individual-exposure visits, a cabinet blocked off one-third of the room (Fig. 1). Four cameras, positioned discreetly around the room, captured everything that occurred in the room and allowed for subsequent coding.

Video stimuli were presented to the infants using custom-designed software that activated one 20-s video when the screen was touched. Additional touches to the screen while a video was playing had no effect. As soon as the video clip ended, subsequent touches triggered a new video. The software was designed to randomize video clips within a given set (i.e., within a given speaker). If the screen was not touched for a period of 90 s, the program automatically triggered a "wake-up clip," a 5-s clip designed to recapture children's interest in the screen. Importantly, the wake-up clips did not contain any of the target Mandarin sounds.

Finally, all infants wore LENA recorders during their exposure visits. LENA devices are designed to capture the naturalistic language environment of children. The LENA recorder can store up to 16 h of digitally recorded sound and can be snapped into a pocket on the front of a child's vest. This allows the recorder to unobtrusively capture all language in the child's environment. The recordings can be subsequently downloaded and analyzed by LENA software. The LENA software provides several automated measures of the language content, but for the purposes of the present study, all language measures were coded manually.

Stimuli. The 20-s videos used during the exposure visits were extracted from the videos used in the video-only condition of Kuhl et al. (2). Videos showed native Mandarin speakers playing with toys, singing nursery rhymes, and reading books. Clips of each of the four native Mandarin speakers were arranged into sets such that in a given visit, children saw one of three sets of clips from one particular speaker. Video clips were selected to contain a high volume of the target contrasts, an alveolo-palatal affricate (/tɕʰi /) and an alveolo-palatal fricative (/ɕi/) Mandarin Chinese consonant contrast that is not phonemic in English or Spanish. The syllables were 375 ms in duration; had identical steady-state vowel formant frequencies of 293, 2,274, 3,186, and 3,755 Hz, respectively; bandwidths of 80, 90, 150, and 350 Hz, respectively; and a fundamental frequency of 120 Hz (high-flat tone, tone 1 in Mandarin). The syllables differed only in the point of maximum rise in amplitude during the initial 130-ms frication portion. The affricate consonant had a fast amplitude rise, with maximum amplitude occurring at 30 ms; the fricative consonant had a slower amplitude rise time, with maximum amplitude occurring at 100 ms. Tokens were equalized in RMS amplitude and played to infants at a comfortable listening level of 67 dBA (2, 60).

In addition to the exposure clips, 5-s videos were extracted to serve as wake-up clips. Critically, wake-up clips did not contain any instances of the target contrast.

Procedure.

Exposure visits. All 12 exposure visits occurred within 4 wk of an infant enrolling in the study. Infants were randomly assigned to individual exposure sessions or paired exposure sessions and infants only experienced one type of exposure during the study.

Individual exposure sessions. When caregivers and infants arrived, the experimenter invited them into the test room and outfitted infants with a LENA recorder and vest. Parents sat in the back of the experimental room and were instructed to limit interactions with their child.

Upon leaving the room, the experimenter retreated to the control room. The experimenter began recording the video from the four cameras in the exposure room and started the computer program. When the program began, the exposure screen changed from black to a still photo of a baby's face. At that point, any time the infant touched the screen, a 20-s video clip played. Touches to the screen during videos were recorded by the touchscreen, but they did not interrupt the video clip. Only when the clip ended would subsequent touches trigger a new video. During the course of the study, no infants had to be shown how to use the touchscreen; all infants figured it out.

Sessions were ended for one of three reasons, whichever came first: (*i*) the infant watched all of the videos in a given set and the program stopped, meaning that infants received at least 20 min of exposure through 20-s clips; (*ii*) the infant became fussy beyond recovery, as determined either by the experimenter or the parent; or (*iii*) two wake-up clips played in a row, meaning that the infant had not touched the screen in 3 min.

At the end of the appointment, caregivers completed a Caregiver Perception Rating form that asked caregivers to rate their perceptions of their child's interest and activity level during the session.

Paired-exposure sessions. Infants in the paired-exposure condition were always partnered with one other child of roughly the same age. Only two babies were in the session at a given time, although infants were not always partnered with the same baby. Infants were intentionally paired with at least three other infants over the course of the 12 exposure visits. The number of partners a given infant was paired with varied based on schedule, family availability, and so forth. At the beginning of paired-exposure sessions, the experimenter waited in the lobby for both families to arrive before going to the experimental space. Once there, the experimenter gave LENA recorders and vests to both infants. Different colors of vests were chosen for partner infants and the color of each child's vest was noted for ease of video coding later. As in the individual-exposure sessions, both caregivers were asked to sit toward the back of the room and to limit their interactions with their baby.

Once in the control room, the experimenter began recording from the exposure room cameras and started the computer program. As with individual sessions, clips played when infants touched the screen. The discontinuation criteria for the paired sessions were identical to that of the individual sessions. Again, both caregivers completed Caregiver Perception Rating forms.

Test visits. After completing all 12 exposure sessions, infants returned to the laboratory for three test sessions: two sessions of a conditioned head turn task and one session for an ERP test. Both of the conditioned head-turn sessions always preceded the ERP test. All test sessions were completed within 2 wk of the final exposure visit.

Conditioned head-turn task. Phonemic learning was assessed in a conditioned head-turn procedure in which infants were trained and then tested on their ability to turn their head toward a loudspeaker when they detect a target phoneme interspersed among the background, or standard sounds (58). In the present study, the alveolo-palatal fricative (/ɕi/) served as the background sound and the alveolo-palatal affricate (/tɕʰi /) was the target sound. During change trials, the background sound changed to the target sound for a 6-s period. When infants turned their head during this period, their head turn was reinforced with a 5-s presentation of a noise-making toy (a monkey playing cymbals or a bear playing a drum) and was recorded as a "hit." Failure to turn toward the toy during a change trial was coded as a "miss." During control trials, the background sound played consistently. Head turns during this period were not reinforced and were coded as "false alarms." Failure to turn during a control trial was recorded as a "correct rejection."

The first of an infant's two visits for the head-turn task was for conditioning. Infants were tested during the second visit. In the conditioning phase, infants only heard change trials in which the target sound was played. Initially, the target sound was played 4-dB louder than the background sound, but after infants correctly anticipated the toy presentation twice in a row, the volume cue was removed. After three additional, consecutive head turns without the volume cue, infants advanced to the test phase.

The second test visit for the head turn task presented infants with equal numbers of control and change trials. Thirty trials were presented in random order, with the caveat that no more than three of the same kind of trial could be presented consecutively. Infant performance on the test trials was calculated as percent correct (% hit + % correct rejection/2).

An experimenter sitting in an adjacent room to the test room watched a live-feed video of the infant to code the procedure in real time. During conditioning, the real-time codes of the experimenter determine when the infant progresses out of the initial training and when they complete the training phase altogether.

Throughout the conditioned head-turn task, the caregiver and the experimenter who was in the room with the child wore headphones that played music to prevent either adult from influencing the child. The experimenter who controlled the task from a control room also wore headphones. These headphones played the background sound but were programmed to go silent during all trials, regardless of whether they were change trials or control trials. This ensured that the experimenter who coded the task was blind to trial type.

ERP test. Mandarin phonemic learning was also tested using the classic oddball paradigm (24) in which a "standard" syllable is repeated on 85% (850 stimulus repetitions) of the trials, and a "deviant" sound is presented pseudorandomly during the remaining 15% of the trials (150 stimulus repetitions). Specifically, the deviant sound did not occur consecutively, and at least three standard sounds were presented between deviant sounds. The time between the offset of a stimulus and the onset of the next stimulus (interstimulus interval) was 705 ms. The number of trials accepted for the standard sound in the individual exposure session was 311.4 (SD = 91.52) and 100.25 (SD = 26.48) trials for the deviant sound. The number of trials accepted for the standard sound in the paired exposure session was 284.53 (SD = 71.57) and 85.53 (SD = 23.04) trials for the deviant sound. No significant differences were found across exposure sessions for the standard sound, $t(29) = 0.907$, $P > 0.05$, or deviant sound, $t(29) = 1.64$, $P > 0.05$.

Infants were awake and tested inside a sound treated room. The child sat on the caregiver's lap. In front of them, a research assistant entertained the child with quiet toys while a muted movie played on a TV behind the assistant. The caregiver and the research assistant wore headphones with masking music during the testing phase. The electroencephalogram (EEG) was recorded using electro-caps (ECI, Inc.) incorporating 32 preinserted tin inverting electrodes. The EEG was referenced to the left mastoid from Fp1, Fpz, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T3, C3, Cz, C4, T4, CP5, CP1, CP2, CP6, T5, P3, Pz, P4, T6, O1 Oz, O2, and RM in the International 10/20 System. Infant eye-blinks were monitored by recording the electrooculogram from one infraorbital electrode placed on the infant's left cheek. The EEG data were collected in DC mode and it was rereferenced off-line to the right mastoid to obtain a more balanced reference distribution. The EEG was recorded using NeuroScan SynAmps RT amplifiers (24-bit A/D converter) using Scan4.5 software. A 1-ms trigger was time-locked to the presentation of each stimulus to accomplish the ERP averaging process (Stim 2 Neuroscan Compumedics).

The amplitudes for the deviant and standard responses were calculated by averaging the voltage values from two ERP time windows: 150–250 ms and 250–350 ms. The mean-amplitude of the deviant ERP response was compared with the mean-amplitude of the standard ERP response. The deviant vs. standard comparison in the 150- to 250-ms time-window range is referred to in the text as pMMR, while the comparison in the 250- to 350-ms time-window range is referred to as nMMR. We also calculated the difference waveform between deviant and standard ERPs by subtracting the standard ERP response from the deviant ERP response (deviant minus standard). Two time windows of interest were evaluated; 150–250 ms after stimulus onset, which is associated with the pMMR, and 250–350 ms after stimulus onset, which is associated with the nMMR response.

The impedances of all electrodes were kept below 5 KΩ. EEG segments with electrical activity ±150 mV at any electrode site were omitted from the final average. EEG segments of 700 ms with a prestimulus baseline time of 100 ms were selected and averaged offline to obtain the ERPs. Baseline correction was performed in relationship to the prestimulus time. The ERP waveforms were band-pass–filtered off-line (1–40 Hz with 12-dB roll off) using the zero-phase shift-mode function in NeuroScan Edit 4.5. The high- and low-cutoff filters used are reported elsewhere and do not produce attenuation of the ERP waveforms (61). The ERP waveforms in Fig. 2 were band-pass–filtered from 1 to 30 Hz (12-dB roll off) for illustration purposes.

Coding. We coded several characteristics of the exposure sessions, starting with children's motor abilities. Because the exposure sessions allowed children to move around the room, infants with more advanced motor abilities were better able to navigate the room, pull themselves up on the ledge by the touchscreen, and occasionally better able to access the touchscreen itself. Accordingly, infants' motor skills were assessed by an adapted version of the Peabody Developmental Motor Scale, which awards points for different levels of locomotion, such as belly crawling, scooting, crawling, taking steps, and so forth. To adequately capture infants' development over time, coders reviewed the videotapes of three sessions of each participant to code children's motor ability. Of the three sessions coded for each participant, sessions were randomly selected from the beginning sessions (1–4), the middle sessions (5–8), and the final session (9–12), such that one coded video represented each interval. Coders reviewed the videos from the selected sessions and awarded infants points based on the motor behaviors children exhibited during the session. For example, a milestone motor behavior for 7-mo-olds is rolling from back to front. Infants were awarded one point if they demonstrated that they could roll from their back or bottom to stomach or hands/knees (one or both sides). If infants remained on their back, they received a score of 0 on this item. Coders reviewed the videotapes for all motor milestones from 7 mo to 13 mo.

Additional elements of the exposure sessions and infant behavior were coded using ELAN coding software. This software allowed coders to code multiple tiers of information in the same platform. For all elements coded in ELAN, eight exposure sessions were randomly selected for coding. The coding window was further refined to a 5-min window of time within the exposure session. In most cases, coders reviewed minutes 5–10 of the exposure session. This eliminated time at the beginning of the session during which infants

were acclimating to the environment. In cases where exposure sessions lasted less than 10 min, the coding window was moved up as needed to produce a 5-min window, such that coders would review minutes 4–9, 3–8, or 2–7. None of the coded sessions lasted fewer than 7 min. Thus, for each infant, coders reviewed eight 5-min segments of exposure sessions. These segments yielded data on touches to the touchscreen, infant vocalizations, and eye gaze. Detailed descriptions of each of these measures are below.

The first of these elements were infant touches to the touchscreen during exposure sessions. Although the experimental computer automatically recorded touches to the screen and whether the touch triggered a new video clip or occurred during video playback, coders manually matched touches to each infant in the paired-exposure condition. Furthermore, for all infants, touches were defined as intentional (infant looked at the screen and touched it) or unintentional (infant touched the screen with a part of their body other than the arm/hand or touched the screen when they were not attending to it).

Next, the LENA audio recordings, in combination with the video recordings, were used to code infant vocalizations. Infant utterances were coded as either nonspeech or as speech sounds. Nonspeech sounds were defined as a sound produced without a pause that is a laugh or cry (with at least one diaphragm flutter), a click, cough, sneeze, raspberry, burp, or exertion grunt. In contrast, speech sounds encompassed all sounds produced without pause that were not nonspeech sounds. Speech sounds could be coded as babble, whine, coo, growl, or squeal (62–64). Utterances were separated by the "breath-group criterion" or a gap in which one can hear an ingress of breath, or a gap sufficiently long for a breath to occur but during which there is no voiced sound (64).

Finally, infants' eye gaze during the exposure sessions was coded in ELAN. Coders used the four complementary video recordings of each session to determine where infants were looking. For infants in individual-exposure sessions, coders noted looks to the screen and to their own caregiver. For infants in paired-exposure sessions, coders recorded these looking patterns, but also recorded looks to the other baby and looks to the other caregiver. When necessary, coders for all infants were also able to designate a time period as "uncodeable" meaning that none of the four camera angles captured an infant's eye gaze.

ACKNOWLEDGMENTS. We thank Sarah Edmunds for her assistance in data collection and coding. This research was supported by Grant SMA-0835854 to the University of Washington LIFE Center from the National Science Foundation Science of Learning Center Program [to P.K.K. (principal investigator)], and from The Ready Mind Project.

1. Anderson DR, Pempek TA (2005) Television and very young children. *Am Behav Sci* 48: 505–522.
2. Kuhl PK, Tsao FM, Liu HM (2003) Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proc Natl Acad Sci USA* 100:9096–9101.
3. Krcmar M, Grela B, Lin K (2007) Can toddlers learn vocabulary from television? An experimental approach. *Media Psychol* 10:41–63.
4. Robb MB, Richert RA, Wartella EA (2009) Just a talking book? Word learning from watching baby videos. *Br J Dev Psychol* 27:27–45.
5. Roseberry S, Hirsh-Pasek K, Parish-Morris J, Golinkoff RM (2009) Live action: Can young children learn verbs from video? *Child Dev* 80:1360–1375.
6. Roseberry S, Hirsh-Pasek K, Golinkoff RM (2014) Skype me! Socially contingent interactions help toddlers learn language. *Child Dev* 85:956–970.
7. Csibra G (2010) Recognizing communicative intentions in infancy. *Mind Lang* 25: 141–168.
8. Myers LJ, LeWitt RB, Gallo RE, Maselli NM (2016) Baby FaceTime: Can toddlers learn from online video chat? *Dev Sci* 20:e12430.
9. O'Doherty K, et al. (2011) Third-party social interaction and word learning from video. *Child Dev* 82:902–915.
10. Nathanson AI (2001) Mediation of children's television viewing: Working toward conceptual clarity and common understanding. *Communication Yearbook 25*, ed Gudykunst WB (Lawrence Erlbaum Associates, Mahwah, NJ), Vol 25, pp 115–151.
11. Hanna E, Meltzoff AN (1993) Peer imitation by toddlers in laboratory, home, and day-care contexts: Implications for social learning and memory. *Dev Psychol* 29:701–710.
12. Ryalls BO, Gul RE, Ryalls KR (2000) Infant imitation of peer and adult models: Evidence for a peer model advantage. *Merrill-Palmer Q* 46:188–202.
13. Ramenzoni VC, Liszkowski U (2016) The social reach: 8-month-olds reach for unobtainable objects in the presence of another person. *Psychol Sci* 27:1278–1285.
14. Chase CC, Chin DB, Oppezzo MA, Schwartz DL (2009) Teachable agents and the protégé effect: Increasing the effort towards learning. *J Sci Educ Technol* 18:334–352.
15. Garcia-Sierra A, Ramírez-Esparza N, Kuhl PK (2016) Relationships between quantity of language input and brain responses in bilingual and monolingual infants. *Int J Psychophysiol* 110:1–17.
16. Friederici AD, Friedrich M, Christophe A (2007) Brain responses in 4-month-old infants are already language specific. *Curr Biol* 17:1208–1211.
17. Rivera-Gaxiola M, Silva-Pereyra J, Kuhl PK (2005b) Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Dev Sci* 8:162–172.
18. Näätänen R (1992) *Attention and Brain Function* (Lawrence Erlbaum Associates, Hillsdale, NJ).
19. Näätänen R (2001) The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology* 38:1–21.
20. Cheour M, et al. (1998a) Maturation of mismatch negativity in infants. *Int J Psychophysiol* 29:217–226.
21. Cheour M, et al. (1997) The mismatch negativity to changes in speech sounds at the age of three months. *Dev Neuropsychol* 13:167–174.
22. Cheour-Luhtanen M, et al. (1995) Mismatch negativity indicates vowel discrimination in newborns. *Hear Res* 82:53–58.
23. Cheour M, Leppänen PHT, Kraus N (2000) Mismatch negativity (MMN) as a tool for investigating auditory discrimination and sensory memory in infants and children. *Clin Neurophysiol* 111:4–16.
24. Näätänen R, et al. (1997) Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385:432–434.
25. Cheour M, et al. (1998b) Development of language-specific phoneme representations in the infant brain. *Nat Neurosci* 1:351–353.
26. Zhang Y, Kuhl PK, Imada T, Kotani M, Tohkura Y (2005) Effects of language experience: Neural commitment to language-specific auditory patterns. *Neuroimage* 26: 703–720.
27. Baldwin DA (1991) Infants' contribution to the achievement of joint reference. *Child Dev* 62:875–890.
28. Brooks R, Meltzoff AN (2002) The importance of eyes: How infants interpret adult looking behavior. *Dev Psychol* 38:958–966.
29. Baldwin DA (1993) Infants' ability to consult the speaker for clues to word reference. *J Child Lang* 20:395–418.
30. Brooks R, Meltzoff AN (2015) Connecting the dots from infancy to childhood: A longitudinal study connecting gaze following, language, and explicit theory of mind. *J Exp Child Psychol* 130:67–78.
31. Dunham PJ, Dunham F, Curwin A (1993) Joint-attentional states and lexical acquisition at 18 months. *Dev Psychol* 29:827–831.
32. Dehaene-Lambertz G, Baillet S (1998) A phonological representation in the infant brain. *Neuroreport* 9:1885–1888.
33. Dehaene-Lambertz G, Dehaene S (1994) Speed and cerebral correlates of syllable discrimination in infants. *Nature* 370:292–295.
34. Ferjan Ramírez N, Ramírez RR, Clarke M, Taulu S, Kuhl PK (2017) Speech discrimination in 11-month-old bilingual and monolingual infants: A magnetoencephalography study. *Dev Sci* 20:e12427.
35. Lee C-Y, et al. (2012) Mismatch responses to lexical tone, initial consonant, and vowel in Mandarin-speaking preschoolers. *Neuropsychologia* 50:3228–3239.
36. Cheng Y-Y, et al. (2015) Feature-specific transition from positive mismatch response to mismatch negativity in early infancy: Mismatch responses to vowels and initial consonants. *Int J Psychophysiol* 96:84–94.
37. Cheng YY, et al. (2013) The development of mismatch responses to Mandarin lexical tones in early infancy. *Dev Neuropsychol* 38:281–300.
38. Kushnerenko E, et al. (2002) Maturation of the auditory event-related potentials during the first year of life. *Neuroreport* 13:47–51.
39. Leppänen PHT, Eklund KM, Lyytinen H (1997) Event-related brain potentials to change in rapidly presented acoustic stimuli in newborns. *Dev Neuropsychol* 13: 175–204.
40. Maurer U, Bucher K, Brem S, Brandeis D (2003) Development of the automatic mismatch response: From frontal positivity in kindergarten children to the mismatch negativity. *Clin Neurophysiol* 114:808–817.
41. Morr ML, Shafer VL, Kreuzer JA, Kurtzberg D (2002) Maturation of mismatch negativity in typically developing infants and preschool children. *Ear Hear* 23:118–136.
42. Trainor L, et al. (2003) Changes in auditory cortex and the development of mismatch negativity between 2 and 6 months of age. *Int J Psychophysiol* 51:5–15.
43. Gomes H, et al. (2000) Mismatch negativity in children and adults, and effects of an attended task. *Psychophysiology* 37:807–816.
44. Shafer VL, Yu YH, Datta H (2011) The development of English vowel perception in monolingual and bilingual infants: Neurophysiological correlates. *J Phonetics* 39: 527–545.
45. Shafer VL, Yu YH, Garrido-Nag K (2012) Neural mismatch indices of vowel discrimination in monolingually and bilingually exposed infants: Does attention matter? *Neurosci Lett* 526:10–14.
46. Tremblay K, Kraus N, McGee T (1998) The time course of auditory perceptual learning: Neurophysiologic changes during speech sound training. *Neuroreport* 9: 3557–3560.
47. Kuhl PK, et al. (2008) Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philos Trans R Soc Lond B Biol Sci* 363:979–1000.
48. Kuhl PK (2007) Is speech learning 'gated' by the social brain? *Dev Sci* 10:110–120.
49. Walton GM, Cohen GL, Cwir D, Spencer SJ (2012) Mere belonging: The power of social connections. *J Pers Soc Psychol* 102:513–532.
50. Brooks R, Meltzoff AN (2005) The development of gaze following and its relation to language. *Dev Sci* 8:535–543.
51. Hirsh-Pasek K, et al. (2015) Putting education in "educational" apps: Lessons from the science of learning. *Psychol Sci Public Interest* 16:3–34.

52. Bigelow AE, MacLean BK, MacDonald D (1996) Infants' response to live and replay interactions with self and mother. *Merrill-Palmer Q* 42:596–611.
53. Goldstein MH, King AP, West MJ (2003) Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proc Natl Acad Sci USA* 100:8030–8035.
54. Kirkorian HL, Choi K, Pempek TA (2016) Toddlers' word learning from contingent and noncontingent video on touch screens. *Child Dev* 87:405–413.
55. Lauricella AR, Pempek TA, Barr R, Calvert SL (2010) Contingent computer interactions for young children's object retrieval success. *J Appl Dev Psychol* 31:362–369.
56. Lindová J, Špinka M, Nováková L (2015) Decoding of baby calls: Can adult humans identify the eliciting situation from emotional vocalizations of preverbal infants? *PLoS One* 10:e0124317.
57. Nunnally JC, Lemond LC (1973) Exploratory behavior and human development. *Advances in Child Development and Behavior*, ed Reese HW (Academic, New York), Vol 8.
58. Werker JF, Polka L, Pegg JE (1997) The conditioned head turn procedure as a method for testing infant speech perception. *Infant Child Dev* 3/4:171–178.
59. Kuhl PK (2011) Early language learning and literacy: Neuroscience implications for education. *Mind Brain Educ* 5:128–142.
60. Tsao F-M, Liu H-M, Kuhl PK (2006) Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants. *J Acoust Soc Am* 120: 2285–2294.
61. Sabri M, Campbell KB (2002) The effects of digital filtering on mismatch negativity in wakefulness and slow-wave sleep. *J Sleep Res* 11:123–127.
62. Gros-Loius JG, West MJ, Goldstein MH, King AP (2006) Mothers provide differential feedback for infants' prelinguistic sounds. *Int J Behav Dev* 30:509–516.
63. Oller DK (2000) *The Emergence of the Speech Capacity* (Lawrence Erlbaum Associates, Mahwah, NJ).
64. Oller DK, Lynch MP (1992) Infant vocalizations and innovations in infraphonology: Toward a broader theory of development and disorders. *Phonological Development: Models, Research, Implications*, eds Ferguson C, Menn L, Stoel-Gammon C (York Press, Parkton, MD), pp 509–536.