

Social Interaction in Infants' Learning of Second-Language Phonetics: An Exploration of Brain–Behavior Relations

Barbara T. Conboy

*Institute for Learning & Brain Sciences, University of Washington, Seattle, Washington, and
Department of Communicative Disorders, University of Redlands, Redlands, California*

Rechele Brooks , Andrew N. Meltzoff , and Patricia K. Kuhl

Institute for Learning & Brain Sciences, University of Washington, Seattle, Washington

Infants learn phonetic information from a second language with live-person presentations, but not television or audio-only recordings. To understand the role of social interaction in learning a second language, we examined infants' joint attention with live, Spanish-speaking tutors and used a neural measure of phonetic learning. Infants' eye-gaze behaviors during Spanish sessions at 9.5–10.5 months of age predicted second-language phonetic learning, assessed by an event-related potential measure of Spanish phoneme discrimination at 11 months. These data suggest a powerful role for social interaction at the earliest stages of learning a new language.

Models of early language development have highlighted the role of social factors (e.g., Bloom, 2000; Bloom & Tinker, 2001; Bruner, 1983; Hoff, 2006; Hollich, Hirsh-Pasek, & Golinkoff, 2000; Kuhl, 2007; Tomasello, 2003; Vygotsky, 1962). These models have been supported by empirical work showing that differences in infants' social behaviors affect language development (e.g., Adamson, Bakeman, & Deckner, 2004; Brooks & Meltzoff, 2008; Mundy et al., 2007) and that social settings facilitate language learning, including early vocal development and first words (e.g., Goldstein & Schwade, 2008; Goldstein, King, & West, 2003; Meltzoff, Kuhl, Movellan, & Sejnowski, 2009; Tamis-LeMonda, Bornstein, & Baumwell, 2001). Specifically, infants' word learning has been linked to their ability to jointly attend to the same referents as their communicative partners (Baldwin, 2000; Brooks & Meltzoff, 2015; Carpenter, Nagell, & Tomasello, 1998).

Social interaction also plays a role in learning at the phonetic level of language. When 9-month-old English-learning infants experienced a nonnative language (Mandarin) through live interactions with adults, television, or audio-only presentations, only those infants who experienced the new language through live interactions showed phonetic learning as assessed with behavioral and brain measures (Kuhl, 2011; Kuhl, Tsao, & Liu, 2003). Thus, although infants can learn phonetically from 1- to 2-minute presentations of isolated syllables in the absence of a

social context (e.g., Maye, Werker, & Gerken, 2002; McMurray & Aslin, 2005), learning the phonemes of a new language from natural language experience appears to be boosted by interaction with a social partner (Kuhl et al., 2003).

Neuroscience provides tools for illuminating how social behaviors influence language development at a mechanistic level. In adults, event-related potential (ERP) measures of brain activity yield a negative response (mismatch negativity, MMN) to a phonemic change in infrequent stimuli (deviant sounds) embedded in a string of background sounds (standards), reflecting neural activity associated with discrimination (Näätänen, 2001; Näätänen et al., 1997). Mismatch responses (MMRs) have also been reported in infants for phonemic discrimination (e.g., Cheour et al., 1998). Moreover, predictive relationships between infant MMRs and later language development have been demonstrated (Kuhl et al., 2008; Molfese & Molfese, 1997; Rivera-Gaxiola, Klarman, Garcia-Sierra, & Kuhl, 2005a), and neural measures of phonetic perception have also been employed to show that infants can learn phonemes from live tutors (Conboy & Kuhl, 2011).

Given that measures of brain activity are sensitive to individual differences in infants' speech perception, Conboy and Kuhl (2011) used an ERP brain measure to assess infants' ability to learn phonemes in a new language, namely Spanish. Beginning at 9.5 months of age, monolingual English-learning infants interacted with native Spanish-speaking adult tutors during 12 sessions, each 25 minutes in duration, over a month's time. Infants' neural responses to a Spanish phoneme contrast were measured prior to and after the Spanish exposure sessions. Infants' neural responses to an English contrast were also assessed both prior to and after exposure to Spanish. The results showed that a significant discriminatory response was elicited to the Spanish contrast after infants were exposed to Spanish, but not before exposure. As expected for the native English contrasts, infants showed a significant discriminatory response before and after exposure to Spanish. Thus, 9.5-month-old infants exposed to complex, naturalistic language (Mandarin or Spanish) in a social context exhibit neural signatures of learning at 11 months of age. However, those exposed to the same amount of input (in Mandarin), but via TV or audio-only input, failed to learn.

A "social gating" hypothesis (Kuhl, 2007) has been offered to explain why phonetic learning occurs when infants listen to complex, natural language during live interaction with tutors, but not from passive listening to TV input. According to this view, language learning is strongly heightened in social settings: When an infant hears speech while interacting with an adult, the infant focuses on linguistic input because the adult's social-communicative intentions make the input salient and because humans increase social arousal (Kuhl, 2011). Both social factors enhance language learning. This saliency does not occur with passive TV or audio-only input in the absence of a social context. Thus far, evidence for the social gating hypothesis of phonetic learning has come from manipulating the form of the delivery of language information provided to infants rather than by examining individual infant-level abilities (Roseberry & Kuhl, 2013).

In the current research, we focused on the relationship between individual infants' phonetic learning and their social behaviors during interactions with adults speaking Spanish. Infants in the Conboy and Kuhl (2011) study demonstrated phonetic learning after a month of Spanish exposure at the group level. Here, we measured joint attention in individual infants by conducting video analyses of infants' eye-gaze behaviors subsequent to the tutors' introduction of new toys during the exposure sessions. We predicted that individual differences in the brain indices of Spanish phonetic learning in the infants in the Conboy and Kuhl (2011) study would be linked to

measures of infants' joint attention as indicated by gaze shifts between the social partner (i.e., tutor) and the object of conversation (i.e., a particular toy). Increased joint attention was hypothesized to be associated with higher levels of Spanish phonetic learning, as reflected in the ERP brain measure, consistent with social gating and other social models of early language development.

METHODS

Participants

Twenty-one infants from monolingual English-speaking homes with a history of fewer than four ear infections, minimum gestation of 37 weeks, and birth weight of at least 2.7 kg were recruited at 40 weeks of age through a university-maintained list (Conboy & Kuhl, 2011). Infants were excluded if parents reported prior experience with a second language or concerns about development or hearing. Infants began the exposure sessions at 42 weeks of age ($M = 41.71$ weeks, $SD = 0.36$). Four children were not included in the final analysis because they left the study ($n = 1$), refused to wear an electrode cap for ERP testing ($n = 2$), or had insufficient ERP data ($n = 1$). Seventeen infants (10 girls) completed all scheduled exposure sessions and had sufficient artifact-free ERP data to be included in the analyses.

Procedure

Infants' gaze behaviors were video recorded during the Spanish exposure sessions (scheduled from 42 to 45 weeks of age). The infants were tested at 46 to 47 weeks of age to assess their neural responses to Spanish and English phonemic contrasts.

Spanish Exposure Sessions

Infants attended twelve 25-minute Spanish exposure sessions. During each session, a native Spanish-speaking adult (tutor) read books with pictures (10 min) and described toys (15 min). A parent seated behind the infant kept the infant facing the tutor and remained silent during the exposure session. All infants had sessions in which they interacted with a tutor alone (single-infant sessions) and in which two infants interacted with the tutor (two-infant sessions). All exposure sessions were video recorded using four temporally synchronized cameras mounted above the infants' heads in each corner of the room (2.7 m by 2.7 m).

Each infant attended sessions with tutors who were blind to the study's hypotheses. We were unable to arrange for each infant to interact with each tutor for the same number of sessions due to scheduling constraints (i.e., 12 sessions in a 4-week period); however, each infant interacted with 3 to 5 different tutors across the 12 sessions. This design ensured that each infant received input in a variety of voices and in a variety of tutor interaction styles.

Three sessions (one early, one midpoint, one later visit) were selected for coding from each infant. The three sessions for each infant involved different tutors. Two were single-infant sessions and the other was a two-infant session. Here we report the two single-infant sessions and their toy-play period to focus on interactions between infants and tutors (Table 1). During

TABLE 1
Spanish Exposure Sessions Assessed for Joint Attention

<i>Subject #</i>	<i>Session Number</i>	
	<i>Early Session</i>	<i>Late Session</i>
S01	3	11
S02	2	9
S03	2	7
S04	4	8
S05	3	12
S06	3	7
S07	2	12
S08	3	11
S09	5	12
S10	2	12
S11	4	9
S12	1	12
S13	2	10
S14	2	7
S15	5	9
S16	2	5
S17	6	11
Mean	3.00	9.65
<i>SD</i>	1.37	2.23

the toy-play period, tutors introduced toys to the infants from a set of 36 items (Figure 1). The tutor presented an individual toy and then continued to play briefly while talking about it in Spanish before selecting another toy. The toys included plastic and stuffed animals, cars, plastic food items, plastic household items, and dolls.

Social Behavior

Independent coders watched the video recordings of the toy-play period to code infants’ eye-gaze behaviors as a measure of infants’ joint attention with their Spanish-speaking tutors. Coders were not aware of the results from the infants’ ERP analyses. Event sampling was conducted during two 3-min video segments from each play period (12 min coded across two sessions). Each event was defined as the 30-sec window following the presentation of each toy.

The coder determined whether infants demonstrated a gaze shift between their social partner and the object of conversation. Gaze shifts were specifically coded when infants shifted their gaze from the tutor’s face to the presented toy (or vice-versa) one or more times during the 30-sec event. There were four other mutually exclusive categories coded: looking to the tutor’s *face-only*, looking to the *toy-only*, looking *away*, and *unable to code* events. Looking *away* was coded when infants did not look at the tutor or the toys. Events were infrequently deemed *unable to code* (5.04% or 20 of 397 events), because the infant’s gaze was ambiguous (e.g., toy was held too near the tutor’s face). The proportion of toy-presentation events in which infants produced a



FIGURE 1 Toy objects presented by the Spanish-speaking tutors (with ruler included to show size).

gaze shift was calculated for each infant for each session, and averaged across the two sessions for the “gaze-shift proportion.” A “face-only proportion,” “toy-only proportion,” and “look-away proportion” were similarly calculated. Rater agreement was determined by having an independent coder rate one session for each infant. The inter-rater agreement for infants’ gaze behavior was excellent (Cohen’s kappa = .81).

General Attentiveness

An independent coder blind to the social-behavior ratings and to the ERP values was used to assess infants’ general attentiveness. The coder observed the entire session (25 minutes including the time with books and toys) and assigned a rating of 1 to 5 based on the infants’ overall attentiveness (with 5 assigned if the infant was highly attentive to the tutor and 1 assigned if the infant was judged to be very inattentive). The “general attentiveness rating” was the average rating across both sessions. This rating system was used in the Kuhl et al. (2003) study, which found that general attentiveness was unrelated to phonetic learning.

Neural Measures of Phonemic Perception

ERPs in response to English and Spanish speech-sound contrasts were collected from each infant following the Spanish exposure sessions. Sufficient artifact-free ERP data (at least 20 trials per stimulus type) were obtained from 17 infants post-exposure (for further details, see Conboy & Kuhl, 2011).

Stimuli and design. ERPs in response to Spanish and English phoneme contrasts were recorded using the stimuli and the double-oddball paradigm (Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005b). In the double-oddball paradigm, three consonant-vowel syllables with the

following initial consonants were used: a voiced /da/ that is phonemic in Spanish but not in English, a voiceless unaspirated consonant that is phonemic both in Spanish (heard as /ta/) and in English (heard as /da/), and a voiceless aspirated /ta/ that is phonemic in English but not in Spanish. The syllable shared by Spanish and English (the voiceless consonant heard as /ta/ in Spanish and as /da/ in English) was the “standard.” Two “deviants” (voiced /da/ that is phonemic in Spanish but not in English, and voiceless aspirated /ta/ that is phonemic in English but not in Spanish) were randomly presented to simultaneously assess both English phoneme discrimination (/da/ – /ta/) and Spanish phoneme discrimination (/ta/ – /da/).

The syllables were naturally produced by a female Spanish/English bilingual speaker and manipulated to obtain a match in duration ($229.65 \pm .3$ msec) and average root mean square values. The voice onset time (VOT) was 24 msec for the Spanish /da/, 46 msec for the English /ta/, and 12 msec for the sound common to both languages. Previous studies demonstrated that adult native English speakers (Rivera-Gaxiola et al., 2005b) and English-learning monolingual infants (Conboy, Rivera-Gaxiola, Klarman, Aksoylu, & Kuhl, 2005; Conboy, Sommerville, & Kuhl, 2008) behaviorally discriminate the English but not the Spanish contrast. Eleven-month-old English-learning monolingual infants show evidence of neural discrimination in the form of a MMR only for the English contrast (Rivera-Gaxiola et al., 2005b), while 10- to 13-month-old Spanish-learning monolingual infants show evidence of neural discrimination (MMR) for the Spanish contrast (Rivera-Gaxiola et al., 2012). In addition, 10- to 12-month-old infants from English-Spanish bilingual homes show a neural response (MMR) to these stimuli that varies in amplitude as a function of English and Spanish exposure (Garcia-Sierra et al., 2011).

The standard syllable was presented on 80% of the trials. The Spanish deviant and the English deviant were each presented on 10% of the trials. Syllables were presented at a comfortable listening level (69 dBA) from two speakers approximately 1.5 m in front of the infant. Infants heard 700 standards, 100 English deviants, and 100 Spanish deviants (with an interstimulus interval of 700 msec) in quasi-random order with at least three standards between deviants.

ERP recordings. Infants sat on a parent’s lap in a sound-attenuated room during ERP testing, watching an assistant play quietly with toys while a silent child-oriented video was displayed on a monitor behind the assistant. One-minute breaks were inserted after every 2 min of stimulus presentation to provide verbal praise and reduce possible habituation to the stimuli.

The electroencephalogram (EEG) was recorded using 32-channel Electro-Caps with tin electrodes arranged as specified by the International 10–20 system and referenced to the left mastoid (offline re-referencing used averaged activity from left and right mastoid electrodes). Electrode impedances were kept under 10 k Ω . Signals were amplified at a gain of 20,000 within a bandpass of 0.1 to 40 Hz, using an Isolated Bioelectric Amplifier System SC-32/72BA (SA Instrumentation, San Diego, CA). Signals were digitized at 250 Hz and stored for further analysis.

ERP data analyses. All data were processed offline, using epochs of 100 msec pre-stimulus and 924 msec post-stimulus onset. Vertical eye movements were detected by opposite polarity in activity at electrodes located below and just above the left eye (FP1). Trials contaminated by eye movement, excessive muscle activity, or amplifier blocking were rejected based on individualized thresholds for each test session and visual inspection of the data. Finally, individual subject averages were filtered using a low pass filter with a cutoff of 15 Hz. Averaged

ERPs were obtained for English deviants, Spanish deviants, and standards occurring immediately before each deviant.

The N250–450 window was selected to capture effects across contrasts based on visual inspection of individual and grand averages. The MMRs to the English and Spanish phonemic contrasts were calculated using the ERPs elicited by a standard stimulus (occurring on 80% of trials) and those elicited by less frequent deviant stimuli. For each infant, averaged English and Spanish MMR effects were calculated by creating difference waves (subtracting the averaged ERP voltages for the standard from those for the English or Spanish deviant at each electrode site), measuring the peak amplitude between 250 and 450 msec relative to the average of a 100 msec prestimulus baseline, and averaging together the values from the midline electrode sites (Fz, Cz, Pz) and lateral electrode sites (FP1/2, F3/4, FC1/2, C3/4, CP1/2).

RESULTS

As previously reported (Conboy & Kuhl, 2011), the postexposure ERP waveforms showed significant phonemic discriminatory responses to both English and Spanish contrasts (Figure 2) at the group level, consistent with learning-related changes in neural activity for the Spanish contrast and the expected response to the English contrast in monolingual infants from English-speaking families. Table 2 presents descriptive data for individual infant's postexposure MMR to

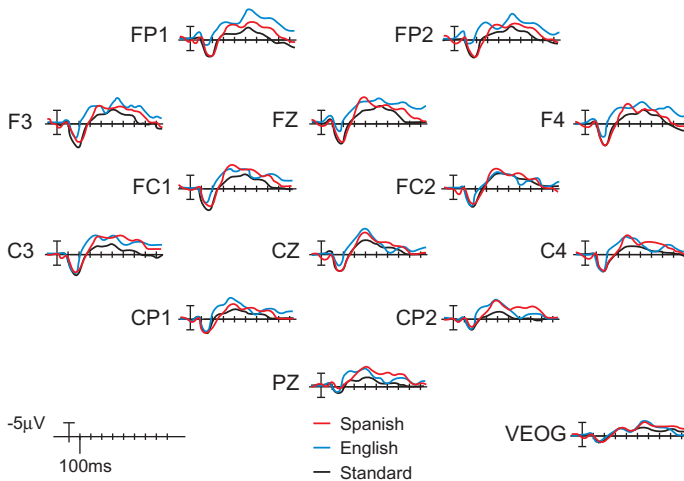


FIGURE 2 Grand-averaged post-exposure event-related potentials (ERPs) to syllables in double-oddball paradigm. Significant mismatch responses were observed between the standard stimulus (voiceless unaspirated consonant that is phonemic both in Spanish as /ta/ and in English as /da/, black line) and the English deviant (voiceless aspirated /ta/ that is phonemic in English but not in Spanish, blue line); and between the standard stimulus (black line) and the Spanish deviant (voiced /da/ that is phonemic in Spanish but not in English, red line). Negative voltages (microvolts) are plotted upward. (Adapted from Conboy & Kuhl, 2011.)

TABLE 2
 Infant Eye-Gaze Behaviors (Proportion) and ERP Mismatch Response (MMR)

<i>Measure</i>	<i>M (SD)</i>	<i>Observed Range</i>
MMR		
Spanish phonemic contrast	-7.06 (4.92)	0.11 to -17.02
English phonemic contrast	-9.40 (5.73)	-1.52 to -19.99
Eye-gaze behaviors		
Gaze-shift proportion	.545 (.126)	.364 to .771
Face-only proportion	.002 (.010)	.000 to .042
Toy-only proportion	.440 (.125)	.229 to .636
Look-away proportion	.015 (.025)	.000 to .077

Note. ERP = event-related potential.

the English and the Spanish phonemic contrasts, and infant’s gaze behaviors during Spanish exposure.

The sessions provided ample opportunities for infants to engage in social behaviors with the Spanish-speaking tutors. There were an average of 11.09 toy-presentation events ($SD = 1.79$) coded in each session, yielding a total of 377 coded toy-presentation events. As a group (Table 2), infants shifted their gaze between their social partner and the object of conversation during a majority of toy-presentation events (gaze-shift proportion $M = .545$), indicating that infants jointly attended to the Spanish-speaking tutor and the toys. Looks to *only* the tutor’s face were exceedingly rare, occurring only once in the entire dataset (*face-only* proportion, $M = .002$). Infants rarely failed to engage with the tutor or the toy (only seven events; *look-away* proportion $M = .015$). The *toy-only* looks accounted for the remaining toy-presentation events (M proportion = .440).

As noted, the *gaze shifts* ($M = .545$) and the *toy-only* looks ($M = .440$) accounted for 99% of the coded behaviors. Examination of the relationship between these two behaviors revealed a highly significant negative Spearman correlation, $r_s(15) = -.99, p < .0001$. The subsequent analyses focused specifically on gaze shifting between the social partner and the object of conversation, which was hypothesized to be associated with higher levels of phonetic learning, as reflected in the ERP brain measure. We acknowledge that any relationships observed with the social behavior of gaze shifting will also exist with the less social *toy-only* looks, but in the opposite direction due to the near perfect inverse relationship between these two behaviors.

The gaze-shift proportion was significantly correlated with the amplitude of the Spanish MMR, $r_s(15) = -.54, p = .026$. Infants who showed a higher proportion of gaze shifts with the tutors showed stronger (more negative) ERP discriminatory responses to the Spanish phoneme contrast (Figure 3). The English MMR was not significantly related to the gaze-shift proportion, $r_s(15) = -.30, p = .24$, indicating that the relationship between brain and behavior we observed was specific to the social situation in which infants were learning the novel second language.

As a check on the specificity of the obtained results, the relationships between infants’ general attentiveness and experimental measures were also examined. The general attentiveness rating did not significantly correlate with the gaze-shift proportion, $r_s(15) = .01 (p = .97)$ or the Spanish MMR, $r_s(15) = -.30 (p = .24)$. Furthermore, the specific session numbers coded for

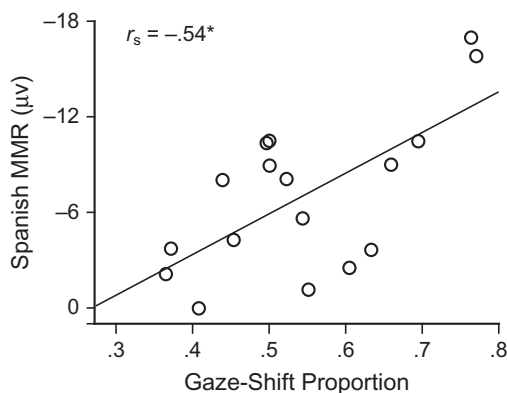


FIGURE 3 Gaze shifts (proportion) during exposure to Spanish significantly correlate with ERP mismatch response (N250–450 difference peak amplitude averaged across 13 midline and lateral electrode sites) to Spanish phonemic contrast. Negative voltages (microvolts) are plotted upward.

individual subjects and the number of toy presentations were not associated with gaze-shift proportion, Spanish MMR, or general attentiveness ($p > .15$).

In sum, infants' joint attention, as assessed by the degree to which they shifted gaze between tutors and toys during the Spanish exposure sessions, significantly predicted the post-exposure strength of the MMR in response to the Spanish contrast. Infants' gaze shifts were not associated with responses to the English MMR, which suggests that the effects of joint attention were specific to new phonetic learning as infants interacted with the Spanish language tutor. Infants' general attentiveness and other session variations did not correlate with phonetic learning.

DISCUSSION

Previous research has demonstrated that social interaction provides support for phonetic perceptual learning in a second language (Kuhl et al., 2003). The current results show, for the first time, that joint attention with language tutors predicts a neural measure of second-language phonetic learning. Infants who interacted socially by actively alternating gaze between the speakers and the objects of conversation (i.e., particular toys) learned more about the linguistic information provided by those speakers. Thus, at the earliest stages of acquiring a language, infants' social behaviors are strongly associated with phonetic learning.

Examining the developmental context of these findings helps in interpreting the correlation between infant behavior and neural indices of learning. Specifically, discrimination of native-language contrasts is well-established by 11 months of age, while discrimination of nonnative contrasts has declined. However, systematic exposure to complex, naturalistic nonnative language in a social context at 9.5 months of age reverses the decline, resulting in nonnative

phonemic discrimination at 11 months (Conboy & Kuhl, 2011; Kuhl et al., 2003). The current work shows that infants' gaze-shift behavior during exposure to a new language at 9.5 to 10.5 months is associated with discrimination of phonemes in that new language at 11 months of age. Further, infants' social behavior correlated with their discriminatory response to Spanish phonemes, but not with their discrimination of native English phonemes. This may indicate a functional connection rather than general language abilities. These findings lead to the working hypothesis that the social behavior of joint attention has powerful effects in learning *new* linguistic information. Future work may determine whether the same social behaviors play a role earlier in development as infants learn native-language phonetic information.

Our results show that individual infants provided with similar social contexts may learn at different levels. In the current study, all infants were presented with the same toys and engaged in the same activities with the same variety of language tutors, yet infants showed different levels of second-language phonetic learning as measured by ERPs. The infants who were more likely to alternate gaze between their Spanish tutors and toys showed more robust phonetic learning, whereas infants more likely to focus on the topic of conversation (i.e., looking only at the toy) evidenced a weaker discriminative response to the Spanish contrast. The association we observed is not explained by the simple sight of a speaker delivering the speech. If simply pairing the speech sound with the image of the speaker were sufficient, infants would learn a second language by watching a video of the tutor, which has been shown to be ineffective in experimentally controlled conditions (Kuhl et al., 2003). Instead, the findings suggest that infants' looking *back and forth* between an actual person and an object of conversation enhances phonetic learning.

Although the current data are correlational and cannot determine causation, the association we observed between infants' social behavior and learning prompts the question: by what mechanism does gaze-shift behavior increase phonetic learning? We offer two possibilities that require additional research.

First, infants who showed gaze shifts between the tutor and the toy may have been learning Spanish word forms, which in turn enhanced phonetic learning. During exposure sessions, infants often watched the tutor intently, closely following the tutor's actions with toys while the tutor spoke. When the infants attended to the social source of information (the tutor) while also attending to the tutor's toys, they may have been exhibiting a form of joint attention, a social behavior linked to word learning (Adamson et al., 2004; Baldwin, 2000; Brooks & Meltzoff, 2008, 2015; Carpenter et al., 1998; Mundy et al., 2007). Learning word forms would be expected to enhance learning of the phonetic units that comprise those word forms (see Yeung & Werker, 2009). It is known that infants at this age segment words from the speech stream and encode the phonetic detail of words in memory (for reviews, see Gervain & Mehler, 2010; Kuhl, 2004). Thus it is possible that the infants exhibiting a higher proportion of gaze shifts were learning nonnative words (aided by the use of joint attention), improving nonnative phonetic perception.

A second possibility is that gaze-shift behaviors reflect infants' increasing cognitive resources and general information processing abilities (Mundy & Jarrold, 2010). Infants' gaze shifts may be part of the attention system and relate to their ability to allocate attention. Both gaze shifts and general attentiveness could indicate how well infants can deploy their attention and learn from the speech stream. However, the ratings of general attentiveness during the exposure sessions were not associated with gaze shifts or with phonetic learning. In contrast, gaze shifts were

significantly associated with phonetic learning, and we believe that this supports a special role for social attention. Future research could include other attentional measures to refine our understanding.

The present data are consistent with the “social gating hypothesis,” which posits that phonetic learning is enhanced by social interaction (Kuhl, 2007). Social interaction enhances language learning in two ways—through increases in motivational factors, such as social arousal, that occur during social interaction (Roseberry & Kuhl, 2013), and through increases in information that are provided by social cues. The current study shows that the information gleaned by infants’ eye-gaze shifts during social interaction is linked to learning.

Given the correlational nature of our findings, it is possible that the increased eye-gaze shifts and differences in neural measures of learning may both be mediated by a third factor not measured in this study. A third factor such as general language processing abilities could underlie the observed relationship; however, the lack of a significant correlation between gaze-shift scores and the English MMR does not support this hypothesis.

Our findings support the contention that social mechanisms are involved in language learning, even during the earliest stages. The association between social behavior and phonetic learning suggests a connection between speech perception and mechanisms of social understanding in humans. This view is buttressed by other work indicating that social interaction plays a role in speech development (e.g., Goldstein & Schwade, 2008; Goldstein et al., 2003).

Future work can further explore the role of infants’ engagement in the social exchange by tracking the point of gaze within a speaker’s face as infants shift their gaze between the speaker and the focal object. Previous eye tracking studies provide evidence that infants focus on the speaker’s mouth (more than the eyes) to glean audio-visual speech information (Lewkowicz & Hansen-Tift, 2012), and that bilingual infants use visual cues from the mouth later in development than monolingual infants (Sebastián-Gallés, Albareda-Castellot, Weikum, & Werker, 2012; Weikum et al., 2007). There is also evidence that infants follow the eyes to gather information about the referent of the speaker’s verbal label (e.g., Baldwin, 2000; Brooks & Meltzoff, 2005). In the current study, we did not have the means to precisely determine whether infants were gazing at their Spanish tutors’ eyes versus mouths. However, our results are the first to link individual differences in brain measures of phonetic learning to infants’ gaze shifts as they alternate gaze between the adult speaker and the object of conversation.

To summarize, the current study revealed a significant relationship between the degree to which infants shifted their gaze between a tutor’s face and the conversation topic and the degree to which infants showed phonetic learning as measured through neural measures. This brain–behavior correlation is statistically significant and theoretically grounded. We argue that social contexts provide important information that is either non-existent or greatly reduced in non-social situations, such as the passive video viewing or the auditory-only presentations that fail to produce phonetic learning (Kuhl et al., 2003). The effects of social interaction on language learning may be multiple and complex. Further research is needed for understanding the full range of factors, both infant-level and situational, that support language learning in social settings as measured both at the neural and behavioral levels.

ACKNOWLEDGMENTS

The authors thank the infants and their families who participated, D. Padden, A. Bosseler, S. Wright, L. Klarman, M. Taylor, and L. Akiyama for their assistance. We are especially grateful to M. Rivera-Gaxiola for providing the ERP paradigm.

FUNDING

Supported by the NSF LIFE Center (SBE-0354453), NIH HD050033 (BTC), and NIH HD37954 (PKK).

ORCID

Rechele Brooks  <http://orcid.org/0000-0002-3798-3448>

Andrew Meltzoff  <http://orcid.org/0000-0001-8683-0547>

REFERENCES

- Adamson, L. B., Bakeman, R., & Deckner, D. F. (2004). The development of symbol-infused joint engagement. *Child Development, 75*, 1171–1187.
- Baldwin, D. A. (2000). Interpersonal understanding fuels knowledge acquisition. *Current Directions in Psychological Science, 9*, 40–45.
- Bloom, L., & Tinker, E. (2001). The intentionality model of language acquisition. *Monographs of the Society for Research in Child Development, 66*(4, Serial No. 267), 1–101.
- Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.
- Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science, 8*, 535–543.
- Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language, 35*, 207–220.
- Brooks, R., & Meltzoff, A. N. (2015). Connecting the dots from infancy to childhood: A longitudinal study connecting gaze following, language, and explicit theory of mind. *Journal of Experimental Child Psychology, 139*, 67–78.
- Bruner, J. (1983). *Child's talk: Learning to use language*. New York, NY: Norton.
- Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4, Serial No. 255), 1–174.
- Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Neuroscience, 1*, 351–353.
- Conboy, B. T., & Kuhl, P. K. (2011). Impact of second-language experience in infancy: Brain measures of first- and second-language speech perception. *Developmental Science, 14*, 242–248.
- Conboy, B. T., Rivera-Gaxiola, M., Klarman, L., Aksoylu, E., & Kuhl, P. K. (2005). Associations between native and nonnative speech sound discrimination and language development at the end of the first year. In A. Brugos, M. R. Clark-Cotton, & S. Ha (Eds.), *Proceedings supplement to the 29th Boston University conference on language development* (pp. 1–11). Boston, MA: Cascadilla Press.
- Conboy, B. T., Sommerville, J. A., & Kuhl, P. K. (2008). Cognitive control factors in speech perception at 11 months. *Developmental Psychology, 44*, 1505–1512.

- Garcia-Sierra, A., Rivera-Gaxiola, M., Percaccio, C. R., Conboy, B. T., Romo, H., Klarman, L., . . . Kuhl, P. K. (2011). Bilingual language learning: An ERP study relating early brain responses to speech, language input, and later word production. *Journal of Phonetics*, *39*, 546–557.
- Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology*, *61*, 191–218.
- Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, *100*, 8030–8035.
- Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*, 515–523.
- Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*, 55–88.
- Hollich, G. J., Hirsh-Pasek, K., & Golinkoff, R. M. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*, *65*(3, Serial No. 262), 1–135.
- Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*, 831–843.
- Kuhl, P. K. (2007). Is speech learning “gated” by the social brain? *Developmental Science*, *10*, 110–120.
- Kuhl, P. K. (2011). Social mechanisms in early language acquisition: Understanding integrated brain systems supporting language. In J. Decety & J. Cacioppo (Eds.), *The Oxford handbook of social neuroscience* (pp. 649–667). Oxford, UK: Oxford University Press.
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory, expanded (NLM-e). *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *363*, 979–1000.
- Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*, 9096–9101.
- Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, *109*, 1431–1436.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101–B111.
- McMurray, B., & Aslin, R. N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition*, *95*, B15–B26.
- Meltzoff, A. N., Kuhl, P. K., Movellan, J., & Sejnowski, T. J. (2009). Foundations for a new science of learning. *Science*, *325*, 284–288.
- Molfese, D. L., & Molfese, V. J. (1997). Discrimination of language skills at five years of age using event-related potentials recorded at birth. *Developmental Neuropsychology*, *13*, 135–156.
- Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., & Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child Development*, *78*, 938–954.
- Mundy, P., & Jarrold, W. (2010). Infant joint attention, neural networks and social cognition. *Neural Networks*, *23*, 985–997.
- 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.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., . . . Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*, 432–434.
- Rivera-Gaxiola, M., Garcia-Sierra, A., Lara-Ayala, L., Cadena, C., Jackson-Maldonado, D., & Kuhl, P. K. (2012). Event-related potentials to an English/Spanish syllabic contrast in Mexican 10–13 month-old infants. *ISRN Neurology*, *2012*, 1–9.
- Rivera-Gaxiola, M., Klarman, L., Garcia-Sierra, A., & Kuhl, P. K. (2005a). Neural patterns to speech and vocabulary growth in American infants. *NeuroReport*, *16*, 495–498.
- Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005b). Brain potentials to native- and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental Science*, *8*, 162–172.
- Roseberry, S., & Kuhl, P. K. (2013). Two are better than one: Infant language learning in the presence of peers. In C. E. Sims (Chairperson), *Language learning and screen media: Investigating the nature of and support for children's learning from video*. Invited symposium conducted at the meeting of the Society for Research in Child Development, Seattle, WA.
- Sebastián-Gallés, N., Albareda-Castellot, B., Weikum, W. M., & Werker, J. F. (2012). A bilingual advantage in visual language discrimination in infancy. *Psychological Science*, *23*, 994–999.

- Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development, 72*, 748–767.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Vygotsky, L. S. (1962). *Thought and language* (E. Hanfmann & G. Vakar, Translators). Cambridge, MA: MIT Press. (Original work published 1934)
- Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007) Visual language discrimination in infancy. *Science, 316*, 1159.
- Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-old infants use distinct objects as cues to categorize speech information. *Cognition, 113*, 234–243.