A Role for Social Interaction in Infants’ Learning of Second-Language Phonetics

Barbara T. Conboy¹², Rechele Brooks,¹ Andrew N. Meltzoff² & Patricia K. Kuhl¹

¹ Institute for Learning & Brain Sciences, University of Washington, USA

² University of Redlands, USA
Abstract

Infants’ learning of phonetic information from ongoing, complex, natural input in a second language has only been demonstrated with live-person presentations, but not television or audio-only recordings (Kuhl, Tsao, & Liu, 2003). To understand the role of social interaction in learning from exposure to a second language, we examined individual differences in infants’ joint engagement with tutors during Spanish-language sessions with live tutors. Infants’ eye gaze behaviors during these sessions at 9.5 to 10.5 months of age were shown to predict individual variation in second-language phonetic learning as assessed by an event-related potential (ERP) measure of Spanish phoneme discrimination following exposure (at 11 months of age). These data suggest a powerful role for social interaction at the earliest stages of learning a new language.

*Keywords:* infant development, language development, speech perception, social interaction, second-language learning, event-related potentials (ERPs)
A Role for Social Interaction in Infants’ Learning of Second-Language Phonetics

Many models of early language development have emphasized the role of social factors in infants’ communicative development. In particular, social interactionist models posit that early language learning is the product of bidirectional influences between infants and adults (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 data showing that social settings provide necessary contexts for learning, and also that learning varies according to differences in infants’ social behaviors. For example, caregiver responsiveness has been linked to variability in the timing of early language milestones, such as the onset of the first word, first fifty words, and combinatorial speech (e.g., Tamis-Lemonda, Bornstein, & Baumwell, 2001), and even to infants’ earlier vocal development (Goldstein, King, & West, 2003; Goldstein & Schwade, 2008). At the same time, there are individual differences in infants’ joint engagement with people and objects during social interactions (e.g., Bakeman & Adamson, 1984; Mundy et al., 2007); infants’ word learning skills have been linked to their ability to understand other’s intentions and jointly attend to the same referents as their communicative partners (Baldwin, 1995; Brooks & Meltzoff, 2008; Carpenter, Nagell, & Tomasello, 1998; Mundy et al., 2007).

There is also evidence to suggest that social interaction plays a role in learning at the perceptual phonetic level of language. When 9- to 10-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 language through live interactions showed phonetic perceptual learning, as measured using a behavioral conditioned head-turn procedure (Kuhl, Tsao, & Liu, 2003). Thus, although statistical learning of phonetic
categories from 1- to 2-minute experimental presentations of repeated syllables (Maye, Werker, & Gerken, 2002; McMurray & Aslin, 2005) can occur in the absence of a social context, when presented with complex natural language, infants may require social interaction to learn from exposure to the new language.

The finding that infants can learn phonemes from live social tutors speaking to them in a second language has recently been extended, using a different language (Spanish vs. Mandarin) and different measures of learning (pre- to post-exposure changes in brain activity vs. post-exposure behavioral measures) than in the Kuhl et al. (2003) study. Monolingual English-learning infants were exposed to Spanish by native Spanish-speaking adult tutors in 12 sessions over a month-long period starting when they were 9.5 months of age (Conboy & Kuhl, 2011).

Previous research has indicated that infants show gradual learning of native-language phonetic categories during the second half of the first year, and at approximately 9 months, there is individual variability across infants in speech perception skills that is linked to intrinsic infant-level factors as well as to amounts of input (for reviews, see Kuhl, 2007; Werker & Tees, 2005). Given that measures of brain activity are sensitive to individual differences in infants’ speech perception skills, the Conboy and Kuhl (2011) study used an event-related potential (ERP) brain measure to determine infants’ ability to learn phonemes in a new language. A discriminatory response, the mismatch response (MMR), was elicited to a Spanish phonemic contrast after, but not before, infants were exposed to Spanish during 12 naturalistic sessions. Infants’ response to an English contrast was also tested before and after exposure, and the MMR was elicited at both ages, though it was more robust at the later age (Conboy & Kuhl, 2011). In summary, 9-10 month-old infants exposed to complex, naturalistic input in a second language in a social context, whether Mandarin or Spanish, show phonetic learning in that language. In contrast, those
exposed to the same amount of Mandarin input in the same voices, but via TV or audio-only input, failed to learn the Mandarin contrast.

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 TV input. From this view, language learning is strongly affected by specific factors that are heightened in social settings. For example, when an infant hears speech while interacting with an adult, the infant may keenly focus on that language input because the adult’s communicative intentions have made the input salient, and this in turn could enhance language learning. This linguistic saliency would not be as readily available from TV or audio-only input because it occurs without a social context. Thus far, evidence for the social gating hypothesis of phonetic learning has focused on the social contexts provided to infants rather than on individual infant-level abilities.

In the current research, we focused on individual variation in phonetic learning and its relationship to variation in infants’ social behaviors during the exposure sessions, all of which took place in social contexts. We predicted that, in spite of hearing similar linguistic input in similar social contexts, the individual infants studied by Conboy and Kuhl (2011) would vary in the brain indices of their phonetic learning and this variation would be linked to measures of infants’ joint engagement with language tutors during the exposure sessions. To measure joint engagement, we conducted video analyses of infants’ eye gaze shifts subsequent to the tutors’ introduction of new toys. More frequent joint engagement was hypothesized to be associated with higher levels of 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 4 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. Infants were excluded if parents reported concerns about development or hearing, or prior experience with a second language. Seventeen infants (10 girls) completed all scheduled exposure sessions and had sufficient artifact-free ERP data to be included in the analyses. Infants began the exposure sessions after 2 weeks of pre-testing, at 42 weeks of age, and ended the sessions at 45-46 weeks, at which time post-testing began. The post-exposure ERP testing took place 1-2 weeks after the last exposure session.

Procedures Used During the Spanish Exposure Sessions

Infants attended twelve 25-minute Spanish exposure sessions across 4 - 5 weeks, beginning at the age of 42 weeks. During each session, a native Spanish-speaking adult (tutor) read scripted picture books (10-15 min) and described toys (10-15 min) while a parent seated behind the infant silently kept the infant facing the tutor. All infants had sessions in which they interacted alone with a tutor and in which another infant also participated. As was done in previous research (Kuhl et al., 2003), each infant received sessions with several different tutors who were blind to the study’s hypotheses. All exposure sessions were video-recorded using four temporally synchronized cameras mounted above the infants’ heads in each corner of a 2.7 m by 2.7 m room.

ERP Analysis of Learning from Spanish Exposure

Approximately one week after the last exposure session, ERP testing was conducted and infants’ MMRs to the English and Spanish phonemic contrasts were measured. Mismatch responses were calculated using the ERPs elicited by a standard stimulus (occurring on 80% of
Infant social behavior and second-language phonetic learning

trials) and those elicited by less frequent deviant stimuli. The standard was a syllable beginning with a consonant that occurs in both Spanish and English (perceived as “ta” by Spanish-speaking adults and “da” by English-speaking adults) and the two deviants were syllables beginning with consonants that occur only in one of the languages (prevoiced Spanish “da”; voiceless aspirated English “ta”). The three syllables were recorded in the same female voice (fundamental frequency, 180 Hz) and digitally controlled for duration, intensity, and average RMS power. Brain electrical activity was recorded using 32-channel Electro-Caps. 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 ms relative to the average of a 100 ms pre-stimulus 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 (see Conboy & Kuhl, 2011, for complete details regarding ERP testing and analyses).

Measurement of Infants’ Social Behaviors

To assess infants’ joint engagement in the task, we examined the toy play period of exposure sessions and selected three representative sessions for each infant such that one was an early visit, one a midpoint visit, and one a later visit, and all three were with different tutors. Of those visits, one was a session in which two infants were present; for the present purposes, only the two sessions in which the infant and parent were alone in the room with the tutor (single infant sessions) were analyzed. We chose to exclude the two-infant sessions in order to focus our analyses on the interactions between each infant and a tutor. Given scheduling constraints (12 sessions within a 5-week period), we were unable to ensure that all infants saw the same tutors for the same numbers of sessions; however, each infant interacted with between 3 and 5 different
tutors across the 12 sessions. This design ensured that each infant received input in a variety of voices and from a variety of tutor styles.

The measure of infants’ level of joint engagement with their Spanish-language tutors was based on the infants’ eye-gaze behaviors, which were assessed by independent coders from video recordings of the Spanish-language sessions. Coders were uninformed of the results from the infants’ ERP analyses. The coder analyzed two 3-minute video segments from each of two sessions (12 minutes total) for each infant in which infants interacted with the tutor. The tutor’s presentation of new toys offered an opportunity to assess infants’ joint engagement. Using a 30-s window following the presentation of each toy, the coder determined whether the infant demonstrated a gaze-shift (GS) between the tutor’s face and the toy. In addition, during the 30-s interval, the infant’s gaze was coded as looking only at the tutor’s face (F) or only at the toy (T), both of which were categorized as “non-gaze shifts” (non-GS). Trials were coded as “look aways” (LA) when infants did not look at the tutor or the toys. Trials were deemed “uncodeable” when the infant’s gaze was ambiguous (as when the tutor held the toy at face level or the camera resolution was too poor for the coder to make a judgment); these occurred on 25 or 5.26% of all 475 trials. The proportion of codeable toy presentations on which infants produced a gaze shift (GS proportion score) was recorded for each infant and averaged across the two sessions. A non-GS proportion score was also calculated. Scorer agreement was determined by having an independent coder rate one of the sessions from each infant (Cohen’s kappa = .80).

**Results**

The results showed that infants’ social behaviors during the Spanish-language sessions were strongly associated with the ERP measures of Spanish phonetic learning. The degree of infants’ engagement with tutors during the sessions, as assessed by average GS proportion scores, was
significantly correlated with infants’ Spanish phonetic learning, as assessed by the amplitude of the Spanish MMR, \( r(15) = -0.63, p < .01 \). Infants who showed a higher proportion of joint engagement with tutors showed stronger and more reliable (more negative) ERP discriminatory responses to the Spanish phoneme contrast (see Figure 1). In contrast, the English MMR did not correlate with the GS score \( r(15) = -0.27, p = .29 \), suggesting that higher joint engagement was linked specifically to Spanish phonetic learning. Infants received, on average, 11.35 toy presentations during the coded portions of each of their sessions (SD = 1.95). Infants were rarely disengaged (i.e., LAs occurred on only 10 of 450, or 2.22%, of the toy presentations), thus these scores were not analyzed further. The infants spent, on average, half of the toy presentations shifting their gaze between a tutor and a toy (mean GS proportion score = .54, SD = .11) and less than half of the toy presentations focusing on either a tutor’s face or toy (mean non-GS proportion score = .44, SD = .11). The non-GS scores were due almost exclusively to infants focusing only on the toy; face-only looks occurred on only 3 of the 450 toy presentations.

To further assess whether infants’ general attentiveness during the sessions could account for the findings, an individual coder who was blinded to the joint engagement ratings and to the ERP results watched the same portions of the sessions that were coded in the present results and assigned a rating of 1 to 5 based on the infants’ overall attentiveness (with a score of 5 assigned when the infant was highly attentive to the tutor and a 1 assigned when the infant was judged to be the least attentive). This rating system followed that used in the Kuhl et al. (2003) study, in which general attentiveness was not found to account for phonetic learning. The present results indicated that no significant relationship existed between these ratings and the infants’ Spanish or English MMR scores (\( r(15) = -0.11, p = .68 \) and \( r(15) = .33, p = .25 \), respectively).
In sum, infants’ social engagement, as assessed by the degree to which they shifted gaze between adult language tutors and the toys the tutors talked about during the Spanish exposure sessions predicted the degree to which they subsequently showed evidence of phonetic learning, as assessed by brain measures. Infants’ social engagement was not correlated with responses to English language responses, which suggests that gaze shifts were specific to tuning into the language of the tutor (Spanish) rather than general language ability. Infants’ general attentiveness to the tutor or to the toy did not correlate with phonetic learning.

Discussion

The current results show, for the first time, that infants’ joint engagement with language tutors predicts learning of the phonetic units of language as assessed by neural measures. Previous research had shown that social interaction provides an important context for phonetic perceptual learning (Kuhl et al., 2003). The current results indicate that infants who interacted socially by actively alternating gaze between speakers of a language and a conversational topic (in this case, a toy) also 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.

Our results show that individual infants provided with similar social contexts may learn at different levels and that the level of learning is related to individual differences in their social behaviors. In the present research, all infants were presented with the same toys and received sessions from the same variety of language tutors, yet infants showed different levels of second-language phonetic learning as measured by ERPs. The infants who showed more robust phonetic learning were significantly more likely to alternate gaze between their Spanish tutors and toys, which may indicate that the infants were jointly engaged with their tutors and the topic (toy).
While the current data are correlational, the association we observed between joint engagement and learning prompts the question: how would joint engagement behavior, as assessed in our study, increase phonetic learning? We offer three possibilities that require additional research.

First, infants who showed gaze shifts to the tutor and the toy when new toys were introduced may have been learning Spanish word forms, which in turn enhanced phonetic learning (see Werker & Yeung, 2005). During exposure sessions, infants often watched the tutor intently, closely following the tutor’s actions with toys while the tutor talked. When 9.5 to 10.5 month-old infants attended to the social source of information (the tutor), while also attending to the tutor’s toys, they may have been exhibiting an early form of “joint visual attention,” a social behavior known to be linked to word learning (Bakeman & Adamson, 1984; Baldwin, 1995; Brooks & Meltzoff, 2008; 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 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 who gaze-shifted the most in this study also were the infants who learned the most phonetically because they did so through the encoding of word forms, which was aided by the use of joint visual attention.

A second possibility is that infants’ statistical phonetic learning from exposure to speech was aided by the presence of visual referents, which provided a meaningful communicative context for learning, and infants who were the most able to make use of such information learned the most. Prior research shows that infants learn from the distributional patterns of phonetic units they experience (e.g., Maye et al., 2002; Saffran et al., 1996). In such experiments, infants
Infant social behavior and second-language phonetic learning

provided with brief auditory-only exposure to speech have retained memory for syllables or word forms. Yet, in complex, naturalistic situations, memory for the distributional statistics of speech may be enhanced when new objects are presented and infants gaze shift between the objects and the speaker, because the social information from speaker signals to the infant the importance of the speech cues. This account would not require word-form learning, per se, but rather, phonetic learning within a meaningful communicative situation.

When infants look at a speaker’s face, they could be using that information to glean audio-visual speech information and/or to glean social-communicative information about the referent of the speaker’s communicative intent. In a recent eye-tracking study, both 8 and 12 month-old infants were found to focus more on speakers’ mouths than on their eyes when the speaker spoke a non-native language (Spanish), whereas they focused more on the speakers’ mouths at 8 months and on the eyes at 12 months when the speaker spoke their native language (English) (Lewkowicz & Hansen-Tift, 2012). These results suggest that information from speakers’ mouths are important for infants’ phonetic learning at times when learning is challenging, such as for native-language processing at 8 months and for non-native processing at both 8 and 12 months. In contrast, infants may shift to the speakers’ eyes to use social cues when under less challenging situations, such as when processing the familiar native language at 12 months. Other research has shown that infants retain the use of visual cues from the mouth for a longer period of time when they are raised bilingually than when they are raised monolingually (Sebastián-Gallés et al., 2012; Weikum et al., 2007). In the current research, we did not have the means to precisely determine whether infants were gazing at their Spanish tutors’ eyes versus the tutors’ mouths. However, our results are the first to link individual differences in learning to the degree to which infants look at their tutors’ faces within the first moments of being presented with a
new toy, suggesting that the use of facial cues is important for phonetic learning and leaving open for future research the question of whether focus on the eyes versus the mouth is associated with individual differences in learning.

A third possibility is that the eye gaze behaviors we measured reflected infants’ growing ability to engage with the tutor at the same time as the objects, which reflects infants' increasing cognitive resources and information processing abilities (Mundy & Jarrold, 2010). The ability to attend to multiple sources of information may play a role in both infants’ social behaviors and language learning. Thus, the infants in our study who only focused on the objects presented by tutors learned less than those who could focus on the tutor’s communicative intent as well as the acoustic cues present in the speech directed at them.

The present data, showing that infants who gaze shift when tutors introduce new objects are also better phonetic learners, are consistent with the “social gating hypothesis,” which posits that phonetic learning is enhanced when infants are engaged with communicative partners in social settings, and are able to actively seek and detect informational linguistic cues that support phonetic learning (Kuhl, 2007). The correlational nature of our findings cannot directly address causal factors; the increased joint visual engagement and differences in neural measures of learning may both be mediated by a third factor not measured in this study. One possible third factor might be general language learning ability; however, the lack of a correlation between the English MMR and GS scores does not support this hypothesis. The findings support the contention that social mechanisms are deeply involved in language learning, even during the earliest stages of learning a language when phonetic information is being learned from distributional cues in the input. The association between social behavior and phonetic learning suggests the potential of a more fundamental connection between language learning and the
mechanisms of social understanding in humans. This view is buttressed by studies that also suggest that social interaction plays a role in speech development (e.g., Goldstein et al., 2003; Goldstein & Schwade, 2008).

To summarize, the current study revealed a highly significant relationship between the degree to which infants shifted their gaze between a tutor’s face and the conversation topic (i.e., a toy) and the degree to which infants showed phonetic learning. We argue that social contexts provide important information that is either non-existent or greatly reduced in non-social situations, such as the video or the auditory-only presentations that did not produce phonetic learning in previous research (Kuhl et al., 2003). Infants bring different social skills to these social contexts, and this may influence learning. We expect that 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.
Acknowledgements

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


Figure Captions

*Figure 1.* ERP mismatch response (N250-450 difference peak amplitude averaged across 13 midline and lateral electrode sites, in microvolts) to Spanish phonetic contrast, by proportion of gaze shifts that occurred during individual Spanish sessions.
Figure 1