Developmental Science 10:1 (2007), pp 110–120 DOI: 10.1111/j.1467-7687.2007.00572.x

Is speech learning 'gated' by the social brain?

Patricia K. Kuhl

Institute for Learning and Brain Sciences, University of Washington, USA

Abstract

I advance the hypothesis that the earliest phases of language acquisition – the developmental transition from an initial universal state of language processing to one that is language-specific – requires social interaction. Relating human language learning to a broader set of neurobiological cases of communicative development, I argue that the social brain 'gates' the computational mechanisms involved in human language learning.

Introduction

Research in the last decade has provided some hints on how infants 'crack the speech code' – they possess powerful computational strategies that have been shown to advance early language learning (Kuhl, 2004; Saffran, 2003). But just as we began to embrace this solution, experimental results tempered the computational conclusion. New studies indicate that early speech learning – at least in complex natural settings – is severely limited in the absence of social interaction, suggesting that something more than passive computation is required to explain infants' rapid language advancement in the first year of life.

Theories of social learning have traditionally emphasized the importance of social interaction for language learning (Bruner, 1983; Vygotsky, 1962). Recent data and theory posit that language learning relies on children's appreciation of others' communicative intentions, their sensitivity to joint visual attention, and their desire to imitate (Baldwin, 1995; Brooks & Meltzoff, 2005; Bruner, 1983; Tomasello, 2003a, 2003b; Tomasello & Farrar, 1986). These theories and models have not, however, been formally extended to early speech learning.

This paper lays out the speech development problem, describes findings showing the impact of social factors on speech development, and advances a new hypothesis: Social interaction is *essential* for natural speech learning. The social hypothesis links early speech learning and neurobiology, and sheds light on developmental disabilities such as autism. The hypothesis also relates speech development to recent findings in neuroscience on 'mirror neurons'. A model of speech development is described that incorporates infants' computational and social abilities (Native Language Magnet-Expanded, or NLM-e) (Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola & Nelson, in press).

The learning problem

Speech learning is a 50-year-old puzzle. On the perception side, how do infants discover which of the sounds used in the world's languages distinguish words in their language? The world's languages contain approximately 600 consonants and 200 vowels (Ladefoged, 2001). Each language uses a unique set of about 40 distinct elements, *phonemes*, which change the meaning of a word (e.g. from *bat* to *pat*). But phonemes are actually groups of non-identical sounds, *phonetic units*, which are functionally equivalent in the language. The baby's task is to make some progress in figuring out the composition of the 40-odd phonemic categories before trying to acquire words on which these elementary units depend. Japanese-learning infants have to group the phonetic units *r* and *l* into a single phonemic category (Japanese *r*), whereas English-learning infants must uphold the distinction to separate *rake* from *lake*. Similarly, Spanish-learning infants must distinguish phonetic units critical to Spanish words (*bano* and *pano*), whereas English-learning infants must combine them into a single category (English *b*). If infants were exposed only to the subset of phonetic units that will eventually be used phonemically to differentiate words in their language, the problem would be trivial. But infants are exposed to many more phonetic variants than will be used phonemically. Simple exposure does not explain phonetic learning.

Address for correspondence: Patricia K. Kuhl, Institute for Learning and Brain Sciences, Mailstop 357920, University of Washington, Seattle, WA 98195, USA; e-mail: pkkuhl@u.washington.edu

Learning to produce the sounds that will characterize infants as speakers of their 'mother tongue' is equally challenging, and is not completely mastered until the age of 8 (Ferguson, Menn & Stoel-Gammon, 1992). Yet, by 10 months, differences can be discerned in the babbling of infants raised in different countries (de Boysson-Bardies, 1993), and in the laboratory, vocal imitation can be elicited by 20 weeks (Kuhl & Meltzoff, 1982). The speaking patterns we adopt early in life last a lifetime (Flege, 1991). My colleagues and I have suggested that this kind of indelible learning stems from a linkage between sensory and motor experience: sensory experience with a specific language establishes auditory patterns stored in memory that are unique to that language; these representations guide infants' successive motor approximations until a match is achieved (Kuhl & Meltzoff, 1996). The process resembles that hypothesized in young birds in which a period of sensory learning of the song elements is followed by a period of motor 'practice' during which infant birds reproduce elements until they achieve a song that approximates the stored auditory template (Doupe & Kuhl, 1999).

What enables this kind of learning for speech? On the perception side, the field of speech recognition technology shows that no machine in the world can derive the phonemic inventory of a language from input (Rabiner & Huang, 1993), though models improve when exposed to 'motherese' (de Boer & Kuhl, 2003). The variability in speech input is enormous; Japanese adults produce both English *r*- and *l*-like sounds, exposing Japanese infants to both sounds (Lotto, Sato & Diehl, 2004; Werker, Pons, Dietrich, Kajikawa, Fais & Amano, in press). Similarly, English speakers produce the Spanish *b* and *p*, exposing American infants to both (Abramson & Lisker, 1970). Infants initially can resolve all these phonetic differences (Eimas, 1975; Eimas, Siqueland, Jusczyk & Vigorito, 1971; Lasky, Syrdal-Lasky & Klein, 1975; Werker & Lalonde, 1988). Thus, the developmental challenge is speech sound *categorization*: Infants must learn which phonetic units are combined to form a phonemic category in their language. How do they do it?

Work in the 1980s indicated *when* infant perception changed – between 6 and 12 months of age, infants' perception of nonnative distinctions declines for a variety of nonnative contrasts (Best & McRoberts, 2003; Werker & Tees, 1984). Our work supports this finding: Japanese infants' discrimination of English *r*-*l* declines between 6 and 12 months (Kuhl, Stevens, Hayashi, Deguchi, Kiritani & Iverson, 2006; see also Tsushima, Takizawa, Sasaki, Shiraki, Nishi, Kohno, Menyuk & Best, 1994), and during the same developmental period, American infants' discrimination declines for a Spanish contrast (Rivera-Gaxiola, Silva-Pereyra & Kuhl, 2005) and a Mandarin contrast (Tsao, Liu & Kuhl, in press) not used in English.

But native-language phonetic perception also changes between 6 and 12 months – experience *facilitates* native-language perception. American and Japanese infants' discrimination of *r* and *l* is equivalent at 6 months, but American infants significantly improve by 10 months (Kuhl *et al*., 2006). Performance increases are shown by 10 months in Mandarin-learning and English-learning infants tested on native-language affricate-fricative contrasts (Tsao *et al*., in press; see Sundara, Polka & Genesee, 2006, for a contrast that shows facilitation later). Event-related potentials (ERPs) also show that native-language discrimination, measured neurally, strengthens between 6 and 12 months, both for consonants (Rivera-Gaxiola, Silva-Pereyra *et al*., 2005) and vowels (Cheour, Ceponiene, Lehtokoski, Luuk, Allik, Alho & Naatanen, 1998).

I have argued elsewhere that the increase observed in native-language phonetic perception between 6 and 12 months represents a critical step in the language acquisition process (Kuhl *et al*., 2006; Kuhl, in press). Our recent data show that native-language discrimination between 6 and 7 months predicts the rate of language growth between 11 and 30 months (Conboy, Rivera-Gaxiola, Klarman, Aksoylu & Kuhl, 2005; Kuhl, Conboy, Padden, Nelson & Pruitt, 2005; Kuhl *et al*., in press; Rivera-Gaxiola, Klarman, Garcia-Sierra & Kuhl, 2005; Tsao, Liu & Kuhl, 2004). Intriguingly, at the cusp of phonetic learning, better performance on *native* contrasts predicts more rapid growth in later language abilities, while better performance on *non*native contrasts predicts slower language growth (Kuhl, Conboy *et al*., 2005; Kuhl *et al*., in press). Based on these findings we argue that exposure to language commits the brain's neural circuitry to the properties of native-language speech, and that neural commitment has bi-directional effects – it increases learning for more complex patterns (such as words) that are compatible with the learned phonetic structure, and decreases the acquisition of nonnative patterns that do not match the learned scheme (Kuhl, 2004). In this framework, the theoretical mechanism explaining phonetic perception cannot be a selection process in which innately specified options are maintained (or lost) as a function of experience, but must instead be a process that depends on growth, a process that relies on infants' discovery of native-language phonetic categories (Kuhl, in press).

A computational solution to phonetic learning

Knowing *when* perception transitions from a universal mode to one that is language-specific did not explain *why*. Research in the 1990s produced an exciting new hypothesis about phonetic learning. Studies indicated that infants learn 'statistically' – they are sensitive to the distributional frequencies of the sounds they hear in ambient language, and this alters perception. Two different approaches with consistent findings established this possibility.

Kuhl and colleagues used a cross-cultural design to test the effects of exposure to the distributional properties of two different languages (Kuhl, Williams, Lacerda, Stevens & Lindblom, 1992). They tested 6-month-old American and Swedish infants with prototypical vowel sounds from both languages, ones that represent the central tendencies (modal values) in adults' speech productions. By varying the critical acoustic components of these prototypical vowels in small steps, 32 variants of each prototype were created. Infants in both countries were tested with the English or Swedish prototype vowel and its variants. The hypothesis was that infants would show a *perceptual magnet effect* (PME) for native-language variants; that is, the native prototype would perceptually 'attract' variants, reflecting prototype learning and categorization, similar to that occurring in infant visual categorization (Quinn, in press). The results confirmed this prediction. American infants perceptually grouped the American vowel variants to a greater extent than the Swedish variants; Swedish infants reversed the pattern, perceptually grouping the Swedish variants more than the American stimuli. The results supported the idea that infants are sensitive to the distributional properties of native language input, and that 6 months of natural listening experience is sufficient to alter infants' perception of speech.

Maye and colleagues conducted experiments that directly assessed whether phonetic learning occurs when the distributional properties of speech are manipulated (Maye, Werker & Gerken, 2002). Maye gave infants short-term (2 min) exposure in the laboratory to sounds from an 8-stimulus continuum to examine the effects of distributional differences on perception. She hypothesized that if infants were sensitive to the distributional properties of sounds they experience, exposure to a bimodal distribution should produce different performance than exposure to a unimodal distribution of the same sounds. Infants at 6 and 8 months were exposed to stimuli on the entire continuum, but the bimodal group heard more frequent presentations of stimuli at the ends of the continuum whereas the unimodal group heard more frequent presentations of stimuli from the middle of the continuum. After familiarization, infants in the bimodal group discriminated the endpoint stimuli on the continuum, whereas those in the unimodal group did not. Taken together, the findings show that phonetic learning can be altered by the distributional patterns contained in language input.

Effects of social factors on learning

Statistical learning occurs in the laboratory when infants are exposed to auditory speech information for only a few minutes (Maye *et al.*, 2002; see Saffran, Aslin & Newport, 1996, for word learning), suggesting that it is an automatic process that requires nothing but the appropriate stimulus. But two studies – one involving speech perception learning and the other speech production learning – suggest that social interaction may be necessary for language learning in natural language-learning situations, and that constraints exist on infants' computational abilities.

A foreign-language intervention design was used to compare the efficacy of live social interaction, as opposed to televised or audio-only presentation, as vehicles for learning foreign-language material. The goal of the studies was to test whether infants can learn phonetically from first-time exposure to a natural foreign language at 9 months, and if so, whether learning depends on social factors (Kuhl, Tsao & Liu, 2003) (Figure 1).

In the live sessions, 9-month-old American infants listened to four native speakers of Mandarin during 12 sessions scheduled over 4–5 weeks' time. The foreign-language 'tutors' read books and played with toys, while speaking natural 'motherese' that was unscripted. A control group was also exposed for 12 sessions but heard only English from American native speakers. After infants completed the sessions, they were tested with a Mandarin phonetic contrast that does not occur in English. Infants showed remarkable learning from the live sessions – they performed significantly better on the Mandarin contrast when compared to the control group that heard only English, and in fact, performed equivalently to infants of the same age tested in Taiwan who had listened to the language for 10 months (Kuhl *et al.*, 2003).

Learning was also assessed using ERPs. The Mismatch Negativity (MMN), shown in previous studies to be sensitive to linguistic experience between 6 and 12 months of age (Kuhl *et al.*, in press; Rivera-Gaxiola, Silva-Pereyra *et al.*, 2005), was measured using the standard 'oddball' paradigm in which infants hear a standard stimulus on 85% of the trials and a deviant stimulus on the remaining 15% of the trials. Statistical comparisons revealed significant differences between the standard and deviant waveforms for the Mandarin Exposure group but not for the English control group.

Live exposure to Mandarin resulted in learning that was durable. Infants returned to the laboratory for their behavioral discrimination tests between 2 and 12 days after the final exposure session, with a median of 6 days, and for their ERP measurements between 8 and 33 days following the last exposure session, with a median of 15 days. These delays allowed us to examine whether longer

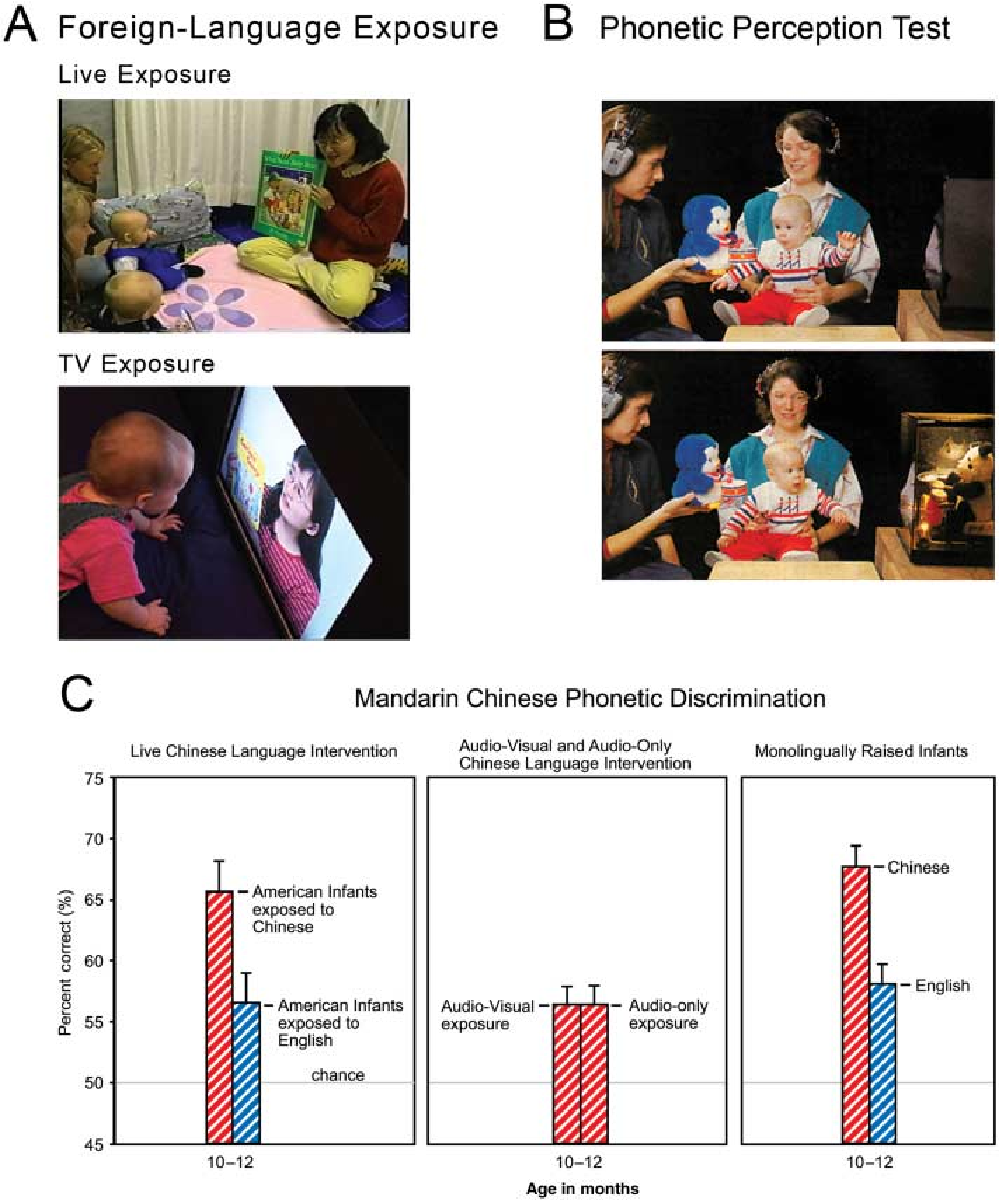

Figure 1 *Infants were exposed to Mandarin via live interactions or via TV or audio-only (A). A control group had live exposure to language for 12 sessions, but heard only English. After exposure, all infants were tested with Mandarin Chinese sounds using a behavioral measure (B). Infants listen to one speech syllable repeated from a loudspeaker on the left while they watch an assistant's toys on the right (B, top). When a second speech syllable is presented, infants are rewarded with a visual stimulus for producing a head-turn toward the loudspeaker (B, bottom). Control trials assess the rate of random head-turn responses. Results indicate learning in the live exposure group, but not in the TV or audio-only groups (C) (From Kuhl* et al.*, 2003, PNAS).*

periods between exposure and test resulted in poorer discrimination. A median-split approach was used to subdivide infants based on the median delay in days between exposure and test. The results indicated no significant differences between discrimination performance for infants above and below the median delay for either the behavioral or brain tests. No 'forgetting' of the Mandarin contrast occurred during the 2 to 33 day delay.

The robustness of learning was surprising and led us to test whether a human being's presence was critical to learning. Two new groups of infants heard the same Mandarin speakers via television or audio-only exposure. The auditory statistical cues available to infants were identical in the televised, audio-only and live settings – if auditory exposure to language automatically induces learning, the presence of a live human being should not be essential.

The results demonstrated an effect of social context. Infants exposed to Mandarin via a standard television or audio-only showed no evidence of learning; their behavioral test scores did not differ from those of infants in the control group who heard no Mandarin whatsoever (Kuhl *et al*., 2003). Thus, the presence of a human being interacting with the infant during language exposure, while not required for simpler statistical-learning tasks (Maye *et al*., 2002; Saffran *et al*., 1996), is critical for learning in our natural language-learning situation in which infants heard, on average, 33,000 Mandarin syllables, from a total of four different talkers, over a 4–5-week period (Kuhl *et al*., 2003).

Social interaction impacts speech production development as well (Bloom, 1975; Bloom & Esposito, 1975; Goldstein, King & West, 2003). In Goldstein *et al*., mothers' responsiveness to their infants' vocalizations was manipulated. After a baseline period of normal interaction, half of the mothers responded immediately to their infants' vocalizations by smiling, moving closer to, and touching their infants; they were the 'contingent condition' (CC) mothers. The other half of the mothers were 'yoked controls' (YC) – their reactions were identical, but timed by the experimenter's instructions to coincide with vocalizations of infants in the CC group. Infants in the CC group produced more vocalizations than infants in the YC group, and their vocalizations were more mature and adult-like. Contingent social interaction appeared to prompt infants to produce the most sophisticated utterances in their repertoires.

What accounts for the impact of social interaction?

Why does social interaction affect early speech learning? We raised two possibilities in our original report. The first was a global mechanism involving infants' *motivation* – the attention and arousal it induces could strongly affect learning. The second was a more specific mechanism involving the *information* content of natural settings – the relations between auditory labels, objects, and speakers' intentions that are available during natural linguistic interaction (Kuhl *et al*., 2003).

Attention and arousal as a mechanism

Attention and arousal affect learning in a wide variety of domains (Posner, 2004). Could they impact learning during exposure to a new language? Infant attention, measured in our studies, was significantly higher in response to the live person than to either inanimate source (Kuhl *et al*., 2003). Attention has been shown to play a role in the distributional learning studies as well. 'High-attender' 10-month-olds learned from bimodal stimulus distributions when 'low-attenders' did not (Yoshida, Pons, Cady & Werker, 2006). And arousal, while not measured in our first tests, appeared to be enhanced. Infants in the live exposure sessions were visibly aroused before the sessions – they watched the door expectantly, and were excited by the tutor's arrival, whereas infants in the non-social conditions did not. Heightened attention and arousal could produce an overall increase in the quantity or quality of the speech information that infants code and remember. Our current studies are testing the hypothesis that individual infants' attention and arousal predict the degree of phoneme and word learning in individual infants in our natural foreign-language learning situation (Conboy & Kuhl, 2006).

Information as a mechanism

We raised a second hypothesis to explain the effectiveness of social interaction – live situations provide specific information that fosters learning (Kuhl *et al*., 2003). During live exposure, tutors focus their visual gaze on pictures in the books or on the toys they talk about, and infants' gaze tends to follow the speaker's gaze (Baldwin, 1995; Brooks & Meltzoff, 2002). Referential information is present in both the live and televised conditions, but it is more difficult to pick up via television, and is totally absent during audio-only presentations. Gaze following is a significant predictor of receptive vocabulary (Baldwin, 1995; Brooks & Meltzoff, 2005; Mundy & Gomes, 1998), and may help infants segment foreign speech. When 9-month-old infants follow a tutor's line of regard in our foreign-language learning situation, the tutor's specific meaningful social cues, such as eye gaze and pointing to an object of reference, might help infants segment words from ongoing speech, thus facilitating phonetic learning of the sounds contained in those words.

Several key developments coincide with the ability to understand reference. By 9 months infants begin to engage in triadic 'person–person–object games' – they systematically combine attention to objects with looks that promote interest from another human, reflecting a 'secondary intersubjectivity' (Trevarthen & Hubley, 1978). Shared perception of communicative intentions, which emerges at around 9 months of age, has been argued to be crucial for the acquisition of language (Akhtar & Tomasello, 1998; Tomasello, 2003a, 2003b). Attending to objects of another person's reference is linked to the infant's growing ability to understand others as intentional agents (Meltzoff, 1995; Tomasello, 2003a). The timing of these social abilities coincides with the beginnings of word comprehension. The suggestion here is that attunement to the communicative intentions of other humans enhances attention to linguistic units at several levels. Attention to the meaning of a communicative act enhances the uptake of units of language present in that act. In our current studies, which involve exposure to Spanish, we are measuring specific interactions between the tutor and the infant to examine whether specific kinds of interactive episodes can be related to learning of either phonemes or words (Conboy & Kuhl, 2006).

What constitutes a social agent?

Our findings raise a more fundamental question: What defines a 'social agent' for infants? Must a social agent involve a human being (with sight, smell, and all other indicators of humanness), or would an inanimate entity, imbued with certain interactive features, induce infant perception of a social being? And if so, could infants learn language from such a socially augmented entity?

Social interaction might be effective *because* it involves other humans, or because features inherent in social settings, such as interactivity and contingency, are critical for learning. Contingency plays a role in human vocalization learning (Bloom, 1975; Bloom & Esposito, 1975; Goldstein *et al*., 2003), and in infant cognition (Watson, 1979, 2005). Interactivity, the reciprocity that is integral in social exchange, could therefore be a key component of speech learning. Infants have a great deal of experience with people whose vocalizations are contingent on their own: Reciprocity in adult–infant language is common as infants alternate their vocalizations with those of an adult (Bloom, Russell & Wassenberg, 1987), and the pervasive use of motherese by adults tends to encourage infant reciprocity (Kuhl & Meltzoff, 1982, 1996).

Whether contingency and interactivity in the absence of a live human would produce learning is an open question. Would infants learn from an interactive TV presentation, one in which the adult tutor was shown on a television but actually performing live from another room so that contingent looking, smiling, and other reciprocal reactions could occur? Could infants learn a new language from a socially interactive robot? Defining what constitutes a social agent for infants is itself of interest, and investigating how the perception of social agency affects learning in young children has both theoretical and practical implications.

Neurobiological connections: communicative learning in animals

Humans are not the only species in which communicative learning is affected by social interaction. Communicative learning in songbirds provides an example. Young zebra finches need visual interaction with a tutor bird to learn song in the laboratory (Eales, 1989). A zebra finch will override its innate preference for conspecific song if a Bengalese finch foster father feeds it, even when adult zebra finch males can be heard nearby (Immelmann, 1969). White-crowned sparrows, which reject the audiotaped songs of alien species, learn the same alien songs when a live tutor sings them (Baptista & Petrinovich, 1986). In barn owls (Brainard & Knudsen, 1998) and white-crowned sparrows (Baptista & Petrinovich, 1986), a richer social environment extends the duration of the sensitive period for learning.

Social contexts also advance song production in birds; male cowbirds respond to the social gestures and displays of females, which affect the rate, quality, and retention of song elements in their repertoires (West & King, 1988), and white-crowned sparrow tutors provide acoustic feedback that affects the repertoires of young birds (Nelson & Marler, 1994).

In birds, social interaction can take various forms. Blindfolded zebra finches that cannot see the tutor but can interact through pecking and grooming learn their songs. And young birds operantly conditioned to present conspecific song to themselves by pressing a key learn the songs they hear (Adret, 1993; Tchernichovski, Mitra, Lints & Nottebohm, 2001). In other words, in birds, interactivity and contingency play critical roles even in the absence of another bird.

Neural underpinnings of social influences on learning

Language evolved to address a need for social communication and evolution may have forged connections between language and the social brain in humans (Adolphs, 2003; Dunbar, 1998; Pulvermuller, 2005). Work on 'mirror neurons' in nonhuman primates indicates a neural link between the self and other; seeing an action and producing it oneself are neurally equivalent in adult monkeys, and this ability plays a role in imitation and social understanding (Meltzoff & Decety, 2003; Rizzolatti, 2005). Research on the development of the neural networks that constitute the 'social brain' is beginning to appear (Johnson, 2005).

A brain area involved in both speech and the mirror system in adults is Broca's area (Pulvermuller, 2005; Rizzolatti & Craighero, 2004). A new imaging technique, magnetoenchephalography (MEG), has now been demonstrated to be feasible for developmental studies of speech perception in infants during the first year of life – MEG allowed us to explore the activation of Broca's area in response to speech (Imada, Zhang, Cheour, Taulu, Ahonen & Kuhl, 2006). Examining brain activation using MEG with infants during social versus nonsocial language experience will allow us to test whether social brain networks are activated differentially in the two conditions.

If social factors 'gate' computational learning, infants would be protected from meaningless calculations – learning would be restricted to signals that derive from live humans rather than other sources (Doupe & Kuhl, 1999; Evans & Marler, 1995; Marler, 1991). Constraints of this kind appear to exist for infant imitation: when infants hear sounds that resemble vowels but could not be produced by human vocal tracts, they fail to imitate (Kuhl, Williams & Meltzoff, 1991), and in action imitation experiments, infants infer and reproduce intentions displayed by humans but not by machines (Meltzoff, 1995). Whether one could induce the perception of social agency in a nonhuman remains to be determined. The current data establish the boundary conditions for language: Exposure to a new language in a live social interaction situation induces remarkable learning in 9-month-old infants, but *no* learning when the exact same language material is presented to infants by a disembodied source.

New model of phonetic learning: NLM-e

We have developed a new model of early speech learning (Kuhl *et al.*, in press), *Native Language Magnet-Expanded* (NLM-e) (Figure 2), which is updated from the original NLM model (Kuhl, 1993, 1994). Of relevance to this discussion is Phase 2 in which *social factors* are argued to play a critical role in phonetic learning.

Phase 2 represents the core of the NLM-e model – in this phase infants become language-specific listeners who produce language-specific vocalizations. Infants' computational skills are depicted in the form of sensitivity to the distributional patterns of vowels and consonants in language input, but sensitivity is affected by social factors. As infants experience language input in natural settings, the brain's neural circuitry is altered to allow easy, automatic detection of the properties of ambient speech. Learning in a social setting is argued to be particularly robust and durable. This process alters perception (see Kuhl, 2004; Kuhl *et al.*, in press, for details).

The model describes neural commitment as influenced by social factors in two ways: first, infants must be attracted to and interested in infant-directed (ID) speech to learn. Infants typically prefer infant-directed (ID) over adult-directed (AD) speech when given an auditory choice in the laboratory (Fernald, 1985; Fernald & Kuhl, 1987); even newborns prefer speech when pitted against nonspeech signals (Vouloumanos & Werker, 2004). However, children with autism spectrum disorder (ASD) do not prefer ID speech. In one study, toddlers with ASD, typical development and developmental delay were allowed to choose between 5-sec 'motherese' speech snippets and nonspeech analogs of the same signals. Only children with ASD showed a preference for the nonspeech analogs (Kuhl, Coffy-Corina, Padden & Dawson, 2005). The degree of preference for the nonspeech signal in childen with ASD predicted the severity of their autism symptoms, the degree of delay in their verbal scores, and the degree of abnormality in their ERP responses to speech syllables (Kuhl, Coffey-Corina *et al.*, 2005).

We have known for some time that a complete lack of social interaction has a devastating impact on human language learning. The few instances in which children have been raised in social isolation show that social deprivation has a severe negative impact on language development, to the extent that normal language skill is never acquired (Fromkin, Krashen, Curtiss, Rigler & Rigler, 1974). These data, and the findings on children with autism, couples early language disabilities and social deficits, suggesting that learning depends on a normal social interest in people and the signals they produce.

Social signals enhance learning because they provide enriched information, such as the referential information discussed earlier. Another example is infant-directed (ID) speech, which exaggerates phonetic distinctions (Burnham, Kitamura & Vollmer-Conner, 2002; Kuhl, Andruski, Chistovich, Kozhevnikova, Ryskina, Stolyarova, Sundberg & Lacerda, 1997; Liu, Tsao & Kuhl, in press). Phonetic cues in ID speech are temporally and/or spectrally 'stretched', increasing the acoustic differences between them and making them more distinct. Greater exaggeration of phonetic differences by mothers is associated with greater learning in infants; infants whose mothers show greater

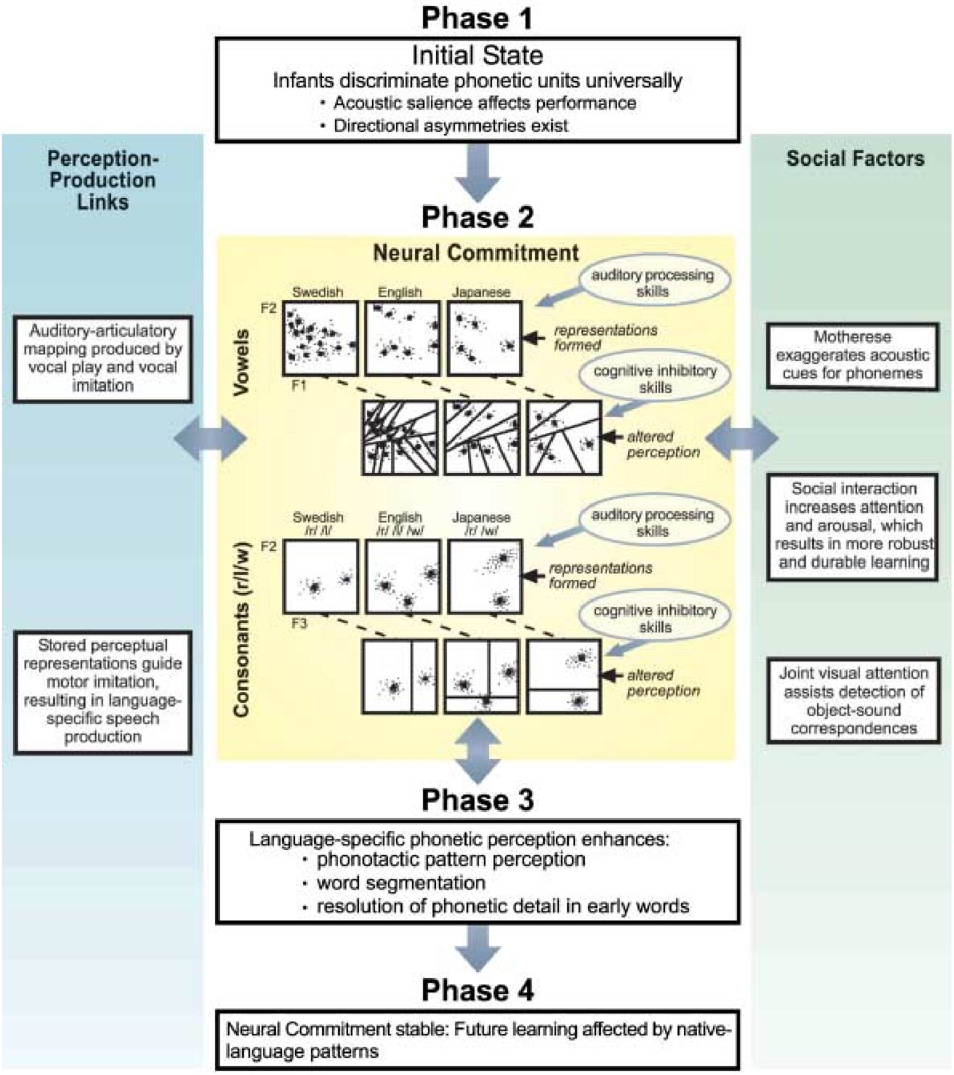

Figure 2 *NLM-e is shown in 4 Phases (see Kuhl* et al., *in press, for full description). In Phase 2, the representations formed by native-language input for vowels and consonants are drawn to reflect existing data for Swedish (Fant, 1973; Lacerda, in preparation), English (Dalston, 1975; Flege* et al., *1995; Hillenbrand* et al., *1995) and Japanese (Iverson* et al., *2003; Lotto* et al., *2004). (From Kuhl* et al., *in press.)*

exaggeration in speech show correspondingly better speech discrimination performance (Liu, Kuhl & Tsao, 2003).

Many questions remain about the impact of social interaction on natural speech and language learning. Whether the underlying mechanism is the increased motivation or the enriched information that social settings provide, or both, remains to be determined. The idea that social interaction is integral to language learning is admittedly not new; however, previous data and theorizing have not tied early speech learning to social factors. Linking early speech acquisition to social cognition places it alongside language acquisition in tapping humans' fundamental social nature. Moreover, it places human communication solidly within the neurobiology of communication learning in nonhuman animals. It is hoped that the ideas and research described here will foster studies that may either enhance or challenge the social hypothesis for speech.

Acknowledgements

The research reported here is supported by an NSF Science of Learning Center grant to the University of Washington's LIFE Center and a grant to PKK from the National Institutes of Health (HD37954).

References

Abramson, A.S., & Lisker, L. (1970). Discriminability along the voicing continuum: cross-language tests. *Proceedings of the Sixth International Conference of Phonetic Sciences in Prague* (pp. 569–573). Prague: Academia.

Adolphs, R. (2003). Cognitive neurosciences of human social behavior. *Nature Reviews Neuroscience*, **4**, 165–178.

Adret, P. (1993). Operant conditioning, song learning and imprinting to taped song in the zebra finch. *Animal Behavior*, **46**, 149–159.

Akhtar, N., & Tomasello, M. (1998). Intersubjectivity in early language learning and use. In S. Bråten (Ed.), *Intersubjective communication and emotion in early ontogeny* (pp. 316–335). Cambridge: Cambridge University Press.

Baldwin, D.A. (1995). Understanding the link between joint attention and language. In C. Moore & P.J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 131–158). Hillsdale, NJ: Lawrence Erlbaum Associates.

Baptista, L.F., & Petrinovich, L. (1986). Song development in the white-crowned sparrow: social factors and sex differences. *Animal Behavior*, **34**, 1359–1371.

Best, C., & McRoberts, G.W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*, **46**, 183–216.

Bloom, K. (1975). Social elicitation of infant vocal behavior. *Journal of Experimental Child Psychology*, **20**, 51–58.

Bloom, K., & Esposito, A. (1975). Social conditioning and its proper control procedures. *Journal of Experimental Child Psychology*, **19**, 209–222.

Bloom, K., Russell, A., & Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *Journal of Child Language*, **14**, 211–227.

Brainard, M.S., & Knudsen, E.I. (1998). Sensitive periods for visual calibration of the auditory space map in the barn owl optic tectum. *The Journal of Neuroscience*, **18**, 3929–3942.

Brooks, R., & Meltzoff, A.N. (2002). The importance of eyes: how infants interpret adult looking behavior. *Developmental Psychology*, **38**, 958–966.

Brooks, R., & Meltzoff, A.N. (2005). The development of gaze following and its relation to language. *Developmental Science*, **8**, 535–543.

Bruner, J. (1983). *Child's talk: Learning to use language*. New York: W.W. Norton.

Burnham, D., Kitamura, C., & Vollmer-Conner, U. (2002). What's new pussycat? On talking to babies and animals. *Science*, **296**, 1435–1435.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Naatanen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, **1**, 351–353.

Conboy, B., Rivera-Gaxiola, M., Klarman, L., Aksoylu, E., & Kuhl, P.K. (2005). Associations between native and non-native speech sound discrimination and language development at the end of the first year. In A. Brugos, M.R. Clark-Cotton, & S. Ha (Eds.), *Supplement to the Proceedings of the 29th Boston University Conference on Language Development*; http://www.bu.edu/linguistics/APPLIED/BUCLD/supp29.html

Conboy, B.T., & Kuhl, P.K. (2006). Event-related potentials to Spanish and English syllables in 11-month-old infants after short-term exposure to Spanish. Poster presented at the NSF Science of Learning Centers' Satellite Meeting at the Society for Neuroscience, 2006.

Dalston, R.M. (1975). Acoustic characteristics of English /w,r,l/ spoken correctly by young children and adults. *Journal of the Acoustical Society of America*, **57**, 462–469.

de Boer, B., & Kuhl, P.K. (2003). Investigating the role of infant-directed speech with a computer model. *Acoustic Research Letters Online (ARLO)*, **4**, 129–134.

de Boysson-Bardies, B. (1993). Ontogeny of language-specific syllabic productions. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 353–363). Dordrecht, Netherlands: Kluwer.

Doupe, A.J., & Kuhl, P.K. (1999). Birdsong and human speech: common themes and mechanisms. *Annual Review of Neuroscience*, **22**, 567–631.

Dunbar, R.J.M. (1998). The social brain hypothesis. *Evolutionary Anthropology*, **6**, 178–190.

Eales, L. (1989). The influences of visual and vocal interaction on song learning in zebra finches. *Animal Behavior*, **37**, 507–508.

Eimas, P.D. (1975). Auditory and phonetic coding of the cues for speech: discrimination of the /r–l/ distinction by young infants. *Perception and Psychophysics*, **18**, 341–347.

Eimas, P.D., Siqueland, E.R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, **171**, 303–306.

Evans, C.S., & Marler, P. (1995). Language and animal communication: parallels and contrasts. In H.L. Roitblat & J.-A. Meyer (Eds.), *Comparative approaches to cognitive science: Complex adaptive systems* (pp. 341–382). Cambridge, MA: MIT Press.

Fant, G. (1973). *Speech sounds and features*. Cambridge, MA: MIT Press.

Ferguson, C.A., Menn, L., & Stoel-Gammon, C. (Eds.) (1992). *Phonological development: Models, research, implications*. Timonium, MD: York Press.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, **8**, 181–195.

Fernald, A., & Kuhl, P.K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, **10**, 279–293.

Flege, J.E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, **89**, 395–411.

Flege, J.E., Takagi, N., & Mann, V. (1995). Japanese adults can learn to produce English /r/ and /l/ accurately. *Language and Speech*, **38**, 25–55.

Fromkin, V., Krashen, S., Curtiss, S., Rigler, D., & Rigler, M. (1974). The development of language in Genie: a case of language acquisition beyond the 'critical period'. *Brain and Language*, **1**, 81–107.

Goldstein, M., King, A., & West, M. (2003). Social interaction shapes babbling: testing parallels between birdsong and speech.

Proceedings of the National Academy of Sciences, **100**, 8030–8035.

Hillenbrand, J., Getty, L., Clark, M., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, **97**, 3099–3111.

Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., & Kuhl, P.K. (2006). Infant speech perception activates Broca's area: a developmental magnetoenceohalography study. *NeuroReport*, **17**, 957–962.

Immelmann, K. (1969). Song development in the zebra finch and other estrildid finches. In R. Hinde (Ed.), *Bird vocalizations* (pp. 61–74). London: Cambridge University Press.

Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, **87**, B47–B57.

Johnson, M.H. (2005). The ontogeny of the social brain. In U. Mayr, E. Awh, & S.W. Keele (Eds.), *Developing individuality in the human brain: A tribute to Michael Posner* (pp. 125–140). Washington, DC: APA Press.

Kuhl, P.K. (1993). Early linguistic experience and phonetic perception: implications for theories of developmental speech perception. *Journal of Phonetics*, **21**, 125–139.

Kuhl, P.K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology*, **4**, 812–822.

Kuhl, P.K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, **5**, 831–843.

Kuhl, P.K. (in press). Linking infant speech perception to language acquisition: phonetic learning predicts language growth. In P. McCardle, J. Colombo, & L. Freund (Eds.), *Infant pathways to language: Methods, models, and research directions*. Erlbaum: New York.

Kuhl, P.K., Andruski, J.E., Chistovich, I.A., Chistovich, L.A., Kozhevnikova, E.V., Ryskina, V.L., Stolyarova, E.I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, **277**, 684–686.

Kuhl, P.K., Coffey-Corina, S., Padden, D., & Dawson, G. (2005). Links between social and linguistic processing of speech in preschool children with autism: behavioral and electrophysiological evidence. *Developmental Science*, **8**, 1–12.

Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (in press). Developmental phonetic perception: native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B*.

Kuhl, P.K., Conboy, B.T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: implications for the 'critical period'. *Language Learning and Development*, **1**, 237–264.

Kuhl, P.K., & Meltzoff, A.N. (1982). The bimodal perception of speech in infancy. *Science*, **218**, 1138–1141.

Kuhl, P.K., & Meltzoff, A.N. (1996). Infant vocalizations in response to speech: vocal imitation and developmental change. *Journal of the Acoustical Society of America*, **100**, 2425–2438.

Kuhl, P.K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show facilitation for native language phonetic perception between 6 and 12 months. *Developmental Science*, **9**, 13–21.

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, USA*, **100**, 9096–9101.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, **255**, 606–608.

Kuhl, P.K., Williams, K.A., & Meltzoff, A.N. (1991). Cross-modal speech perception in adults and infants using nonspeech auditory stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, **17**, 829–840.

Lacerda, F. (in preparation). Acoustic analysis of Swedish /r/ and /l/.

Ladefoged, P. (2001). *Vowels and consonants: An introduction to the sounds of language*. Oxford: Blackwell Publishers.

Lasky, R.E., Syrdal-Lasky, A., & Klein, R.E. (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*, **20**, 215–225.

Liu, H.-M., Kuhl, P.K., & Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, **6**, F1–F10.

Liu, H.-M., Tsao, F.-M, & Kuhl, P.K. (in press). Lexical tone exaggeration in Mandarin infant-directed speech. *Developmental Science*.

Lotto, A.J., Sato, M., & Diehl, R. (2004). Mapping the task for the second language learner: the case of Japanese acquisition of /r/ and /l/. In J. Slitka, S. Manuel, & M. Matthies (Eds.), *From sound to sense* (pp. C181–C186). Cambridge, MA: MIT Press.

Marler, P. (1991). The instinct to learn. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition* (pp. 37–66). Hillsdale, NJ: Lawrence Erlbaum Associates.

Maye, J., Werker, J.F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, **82**, B101–B111.

Meltzoff, A.N. (1995). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, **31**, 838–850.

Meltzoff, A.N., & Decety, J. (2003). What imitation tells us about social cognition: a rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society B*, **358**, 491–500.

Mundy, P., & Gomes, A. (1998). Individual differences in joint attention skill development in the second year. *Infant Behavior and Development*, **21**, 469–482.

Nelson, D., & Marler, P. (1994). Selection-based learning in bird song development. *Proceedings of the National Academy of Sciences, USA*, **91**, 10498–10501.

Posner, M.I. (Ed.) (2004). *Cognitive neuroscience of attention*. New York: Guilford Press.

Pulvermuller, F. (2005). Brain mechanisms linking language to action. *Nature Reviews Neuroscience*, **6**, 574–582.

Quinn, P.C. (in press). Categorization. In A. Slater & M. Lewis (Eds.), *Introduction to infant development* (2nd edn.). Oxford: Oxford University Press.

Rabiner, L.R., & Huang, B.H. (1993). *Fundamentals of speech recognition*. Englewood Cliffs, NJ: Prentice Hall.

Rivera-Gaxiola, M., Klarman, L., Garcia-Sierra, A., & Kuhl, P.K. (2005). Neural patterns to speech and vocabulary growth in American infants. *NeuroReport*, **16**, 495–498.

Rivera-Gaxiola, M., Silvia-Pereyra, J., & Kuhl, P.K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental Science*, **8**, 162–172.

Rizzolatti, G. (2005). The mirror neuron system and imitation. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation: From neuroscience to social science – Volume 1: Mechanisms of imitation and imitation in animals* (pp. 55–76). Cambridge, MA: MIT Press.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, **27**, 169–192.

Saffran, J.R. (2003). Statistical language learning: mechanisms and constraints. *Current Directions in Psychological Science*, **12**, 110–114.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month old infants. *Science*, **274**, 1926–1928.

Sundara, M., Polka, L., & Genesee, F. (2006). Language-experience facilitates discrimination of /d-th/ in monolingual and bilingual acquisition of English. *Cognition*, **100**, 369–388.

Tchernichovski, O., Mitra, P., Lints, T., & Nottebohm, F. (2001). Dynamics of the vocal imitation process: how a Zebra Finch learns its song. *Science*, **291**, 2564–2569.

Tomasello, M. (2003a). *Constructing a language*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2003b). The key is social cognition. In D. Gentner & S. Kuczaj (Eds.), *Language and thought* (pp. 47–58). Cambridge, MA: MIT Press.

Tomasello, M., & Farrar, M.J. (1986). Joint attention and early language. *Child Development*, **57**, 1454–1463.

Trevarthen, C., & Hubley, P. (1978). Secondary intersubjectivity: confidence, confiding, and acts of meaning in the first year. In A. Lock (Ed.), *Action, gesture, and symbol* (pp. 183–229). London: Academic Press.

Tsao, F.-M., Lui, H.-M., & Kuhl, P.K. (2004). Speech perception in infancy predicts language development in the second year of life: a longitudinal study. *Child Development*, **75**, 1067–1084.

Tsao, F.-M., Liu, H.-M., & Kuhl, P.K. (in press). Perception of native and nonnative affricate-fricative contrasts: cross language tests on adults and infants. *Journal of the Acoustical Society of America*.

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., Menyuk, P., & Best, C. (1994). Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6–12 months: language-specific developmental changes in speech perception abilities. Paper presented at the 1994 International Conference on Spoken Language, Yokohama, Japan.

Vouloumanos, A., & Werker, J.F. (2004). Tuned to the signal: the privileged status of speech for young infants. *Developmental Science*, **7**, 270–276.

Vygotsky, L.S. (1962). *Thought and language*. Cambridge, MA: MIT Press.

Watson, J.S. (1979). Perception of contingency as a determinant of social responsiveness. In E.B. Thomas (Ed.), *The origins of social responsiveness* (pp. 33–64). New York: Erlbaum.

Watson, J.S. (2005). The elementary nature of purposive behavior: evolving minimal neural structures that display intrinsic intentionality. *Evolutionary Psychology*, **3**, 24–48.

Werker, J.F., & Lalonde, C. (1988). Cross-language speech perception: initial capabilites and developmental change. *Developmental Psychology*, **24**, 672–683.

Werker, J.F., & Tees, R.C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, **7**, 49–63.

Werker, J.F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (in press). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*.

West, M., & King, A. (1988). Female visual displays affect the development of male song in the cowbird. *Nature*, **334**, 244–246.

Yoshida, K.A., Pons, F., Cady, J.C., & Werker, J.F. (2006). Distributional learning and attention in phonological development. Paper presented at International Conference on Infant Studies, Kyoto, Japan, 19–23 June.