CHAPTER 10

Early speech perception

Developing a culturally specific way of listening through social interaction

Barbara T. Conboy and Patricia K. Kuhl
University of Washington

Throughout the first year of life, infants experience dramatic changes in speech sound perception that reflect a move from universal to specific ways of listening appropriate for their language community. In this chapter, we explore the role of social experience in this important transition in language development. Focusing on the phonetic aspects of language acquisition, we ask: what aspects of language experience serve as agents of change in helping infants to become perceptually attuned to other speakers of the language? We begin with a brief summary of the literature on the development of speech perception, which illustrates the importance of language experience during infancy for establishing native-like speech perception abilities (more extensive reviews of infant speech perception research are available from Goodman & Nusbaum 1994; Kuhl 2004; Jusczyk 1997; Werker & Tae 2005). Next, we review studies in which we have applied the "Conditioned Head Turn" technique to investigate the role of language experience in influencing developmental patterns of speech perception. We then review the results of a recent study that suggest that when a new language is introduced towards the end of the first year, infants participate through social interaction in the process of phonetic learning, rather than learning solely through passive listening. Thus, the language experience required for effective phonetic learning has a highly social nature.

We suggest that particular social cues play an important role in heightening infants' attention to relevant language stimuli in such early second language learning situations, and may also be essential for first language phonetic learning. Based on studies of social-cognitive development during the first year and its relationship to early language acquisition, we suggest that the process of attunement to social information and a sharing of perception throughout the first year direct infants' at-
Overview of developmental speech perception research

Three decades of research on infant speech perception have shown that shifts in speech sound perception occurring over the course of the first year are driven by experience with ambient language. Following a landmark study in which categorical perception of speech sounds was discovered in 1-month-old infants (Eimas et al. 1971), researchers began to explore how perception of particular speech contrasts varied as a function of the language spoken to the infant (e.g., Aslin et al. 1981; Eilers, Gavin & Oller 1982; Eilers, Gavin Wilson 1979; Lasky et al. 1975; Streeter 1976; Trehub 1976; Werker et al. 1981). The work of Werker and Tees (1984a) indicated a developmental progression from similar discrimination of native and nonnative contrasts at 6–8 months, to lack of discrimination of the same nonnative contrast at 10–12 months. Further research replicated these results, leading to a now widely cited developmental pattern of speech perception: infants' ability to discriminate a variety of speech sounds occurring across the world's languages is initially unconstrained by the language of their community; this subsequently gives way to language-specific patterns of discrimination as early as 6 months of age for vowels (Kuhl et al. 1992) and by 10–12 months for consonants (Best, McRoberts, & Sitch 1988; Best et al. 1995; Bosch & Sebastián-Gallés 2003; Burns, Werker, & McVie 2003; Conboy et al. 2005; Kuhl et al. 2001; Kuhl, Tsao, & Liu 2003; Pegg & Werker 1997; Werker & Lalande 1988). Recent studies using event-related potentials (ERPs) to measure brain activity have provided additional evidence for changes in speech perception over the first year of life (Cheour et al. 1998; Kuhl et al. 2007; Rivera-Gaxiola, Silva-Pereyra, & Kuhl 2005a).

Several studies have shown that this well-documented reduction in the perception of nonnative phonemes does not reflect a loss of sensory ability due to mere lack of exposure to the sounds of nonnative languages (see Werker 1994). First, adults retain the ability to discriminate some nonnative phoneme contrasts under certain testing procedures (Carney, Widin, & Viermeister 1977; Werker & Logan 1985; Werker & Tees 1984b), and can learn to discriminate many other nonnative phonemes given training (e.g., Jamieson & Morosan 1986, 1989; Logan, Lively, & Pisoni 1991; McClaskey, Pisoni, & Carrell 1983; McClelland, Fiez, & McCandliss 2002; Morosan & Jamieson 1989; Pisoni et al. 1982, 1994; Tees & Werker 1984; Zhang et al. 2005). Second, reduction in the perception of nonnative phonemes is not uniform, but rather appears to be modulated by acoustic salience (Burnham 1986), by the relationship of the nonnative phonemes to phonemic categories in the adult's native language (e.g., Best 1994; Best & McRoberts 2003; McRoberts & Goodell 2001; Best, McRoberts, & Sithole 1988; Best et al. 1995; Guion et al. 2000; Strange et al. 1998), or by acoustic factors and/or phonetic familiarity (Polka 1991, 1992; Polka, Colantoni, & Sundara 2001).

Native language learning drives the development of native-like speech perception

While most previous studies have focused on the decline of nonnative speech perception, we have proposed that native language perception improves over the first year of life, and furthermore, that changes in nonnative speech perception are linked to such sharpening of perception for the native language (Kuhl 2000). Our lab has conducted a series of behavioural studies that indicate that improvement in native language phonetic perception throughout infancy accompanies a reduction in nonnative perception (Kuhl et al. 2005). For these behavioural studies, we have used the "Conditioned Head Turn Procedure" (HT), a widely used method for testing infant speech perception (Eilers et al. 1979; Kuhl 1979, 1985; Polka, Jusczyk, & Rvachew 1995; Werker et al. 1981). In our version of the HT task, infants sit on their parent's lap while an assistant, seated to the right, manipulates silent toys to attract the infant's attention. Infants are trained to turn away from the assistant and toward a loudspeaker on their left when they detect a change from the repeating background sound to the target sound. An experimenter observes the infant on a video monitor in a control room during testing and judges the head turn responses. Correct head turn responses are reinforced with presentation of a mechanical toy (e.g., bear tapping on a drum) next to the loudspeaker. The HT procedure consists of a conditioning phase followed by a test phase (Figure 1). In the conditioning phase, all trials are change trials, allowing the infant to learn the association between the target sound and visual reinforcement. During conditioning, the target sound is initially presented with an intensity cue to draw the infant's attention to the stimulus change. Following two consecutive correct head-turn responses to the target sound in anticipation of the reinforcer, no-intensity cue trials...
are administered until three consecutive correct head-turn responses have been achieved. In the test phase, change (sound change) and control (no sound change) trials occur with equal probability (50%). For change trials, head-turns are scored as "hits" and failure to turn as "misses"; for control trials, head-turns are scored as "false alarms" and failure to turn as "correct rejections." Several measures are taken to control bias: (a) all contingencies and trial selection are under computer control; (b) the parent and assistant wear headphones and listen to music that masks the speech sounds and prevents them from influencing the infants' responses; and (c) the experimenter's headphones, which allow monitoring of the experimental room, are deactivated during trials so that the experimenter cannot hear the stimuli during the trial. In addition, the results are analyzed using signal detection measures that take both hit and false alarm rates into account.

Using the HT technique, along with a parent-report inventory of language development—the MacArthur-Bates Communicative Development Inventory, or CDI (Fenson et al. 1993)—we have shown that individual variation across infants in the attainment of native-like speech perception is linked to advances in other aspects of language development. As early as 6 months of age infants display language-specific ways of perceiving vowel contrasts (Kuhl et al. 1992). Infants who are better at vowel discrimination at 6 months have better language skills throughout the 14 to 30 month period (Tsao, Liu, & Kuhl 2004). For consonants, the shift to language-specific processing takes several months longer (Conboy et al. 2005; Kuhl et al. 2001, 2005). Infants who are more language-specific listeners for consonants at 7 months also tend to have faster growth in language development from 14 to 30 months (Kuhl et al. 2005). In this research, better performance on native language phoneme discrimination was positively correlated with later CDI scores, whereas better performance on nonnative phoneme discrimination was negatively correlated with later CDI scores. Infants who have higher native-language speech discrimination scores relative to their nonnative discrimination scores also have higher concurrent CDI vocabulary scores (Conboy et al. 2005). Studies using event-related potentials have likewise shown that the attainment of native-like speech perception is linked to subsequent language skills (Kuhl et al. 2007; Rivera-Gaxiola et al. 2005b).

These results indicate that the shift to a language-specific way of listening may be considered a developmental milestone that is continuous with subsequent learning in that language. Infants who more quickly learn to tune out phonetic contrasts that are not meaningful for their native language are also more proficient at detecting contrasts that are phonemic in their language. Infants who reach the milestone of native-like speech perception sooner have an advantage in other aspects of language acquisition. Experience with the native language, and the uptake of information from such experience, appears to influence language acquisition on several levels, beginning with the phonetic level. Although these correlational results do not necessarily establish that phonetic learning drives other aspects of language acquisition, they suggest, at a minimum, that there is continuity across domains and that learning from experience may be essential for each (see also Jusczyk 1997; Werker & Curtin 2005).

We have suggested elsewhere (Kuhl et al. 2005, 2007) that both the decline in the perception of nonnative phonemes during the first year, and constraints on learning at later ages, arise from native-language learning that begins during the first year of life. On this view, native-language learning produces dedicated neural networks that code the patterns of native-language speech, resulting in a "warping" of perceptual representations of the acoustic properties of speech sounds (Kuhl 2000; Kuhl 2004; Kuhl et al. 2005). The result of such neural commitment is resistance to learning phonetic features that are in opposition to those of the native language. Although early proposals regarding "critical periods" for language acquisition (e.g., Lenneberg 1967) asserted that a second language could be acquired without a foreign accent anytime until puberty, more recent research suggests that optimal learning of a second language's phonology occurs much earlier than puberty. For example, Fellow and colleagues have shown that learning a second language after approximately 5 years of age leads to more difficulty perceiving particular speech sounds in that language when compared to acquisition that takes place before that age (Flege, Bohn, & Jang 1997; Flege & Eefting 1987; Flege & MacKay 2004; Flege, MacKay, & Meador 1999). For vowels, acquisition as early as 3–4 years of age may still not result in native-like perception (Bosch, Costa, & Sebastián-Galles 2000; Pallier, Bosch, & Sebastián-Galles 1997). Adults may be trained to perceive nonnative phonemic contrasts, but this does not always result in native-like speech perception (Poikka 1991). In the rest of this chapter we explore some of the factors that might be important for phonetic learning in infancy to be successful. We start with the assumption that brain plasticity remains open for second language acquisition throughout infancy, but also suggest that certain environmental conditions are important for learning to ensue.

Phonetic perception in infants exposed to a second language at 9–10 months

Kuhl and colleagues (Kuhl, Tsao, & Liu 2003) applied the HT technique to address the question of how infants who are well within the "critical" or "sensitive" period for second language acquisition learn to perceive speech sound contrasts from a nonnative language. In Experiment 1, two groups of 10–11-month-old infants growing up in Seattle in monolingual English-speaking homes were tested on a phonetic discrimination HT task using the Mandarin alveolar-palatal /\j/ vs. /\\j/ fricative/affricate contrast, two sounds that are not English phonemes. Although English uses a fricative/affricate phonemic contrast ("sh" vs. "ch"), it does so at
Early speech perception

Mandarin Chinese Phonetic Discrimination

A Experiment 1: Live Chinese-language Intervention
B Experiment 2: Audio-Visual and Audio-Only Chinese-language Intervention
C Monolingual Training Experience

Figure 2. (A) Experiment 1. Mandarin Chinese speech discrimination tests conducted on infants after exposure to Mandarin Chinese (left stripes) or American English (right stripes) show significant learning for the Mandarin exposed infants when compared with the English controls. (B) Experiment 2. Mandarin Chinese foreign-language exposure in the absence of a live person (AV or A) shows no learning. (C) Results of the same Mandarin speech discrimination tests on monolingual Mandarin-learning (grey to the left) and English-learning (black to the right) infants. (From 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 (15). Copyright (2003) National Academy of Sciences, USA.)

In Experiment 2, two groups of infants were brought to the lab for twelve 25-minute sessions. The first group (Mandarin audiovisual-exposure) watched DVDs of the same Mandarin speakers who were used in Experiment 1, showing the same toys and books, but there was no live interaction. The second group (Mandarin audio-exposure) listened to the audio channel of the same DVDs, but received no visual input from speakers of the language. Both groups heard the same amount of Mandarin as the live-exposure group from Experiment 1, delivered in the same naturalistic infant-directed speech by the same speakers. These two groups of infants were then tested on the same Mandarin fricative/affricate contrast used in Experiment 1.

The results of the testing were surprising. The Mandarin-audioonly and Mandarin-audiovisual groups both performed similarly to the English-only control group from Experiment 1; all three groups performed similarly to a separate group of monolingual English infants who were tested at the same age but had never been to the laboratory before (Figure 2). These results show that something special happened during the live Mandarin exposure sessions. Passive listening to
Mandarin from a television did not induce learning, even when visual as well as audio cues were available. All three learning conditions were similar in the sense that language input was provided using a natural voice, and the infants were not required to do anything except sit and listen. Yet phonetic learning was not triggered simply by auditory, or even audio-visual, exposure to Mandarin, in contrast to the previous studies that found that 6–8-month-old infants could learn statistically from small amounts of audio-only exposure to language stimuli.

As discussed previously by Kuhl and colleagues (2003), naturalistic exposure to multiple speakers presents infants with a much more complex learning situation than controlled exposure to isolated instances of speech stimuli during an experiment. Learning occurs in both instances. However, learning in the natural exposure experiments requires much more from infants; they must extract phonetic information for the target contrast over a much broader and more variable range of exemplars. For example, infants in the exposure experiments heard between 26,000 and 42,000 (mean = 33,000) syllables over the course of the experiment, spoken by 4 different people with different voices and styles of speaking. In the statistical learning task conducted by Maye et al. (2002), infants heard tokens of 10 computer-synthesized syllables. The infants exposed to Mandarin in the live condition were shown books and toys that they could track visually. The pairing of auditory and visual information in meaningful interactive contexts may have engaged infants in the live-exposure group in a way that did not occur for the infants in the group that simply watched and listened to the same material over a television screen, or the infants who only heard the speakers’ voices but did not have visual cues.

In complex naturalistic situations, social interaction could be a useful mechanism for heightening infants’ attention to relevant linguistic cues in the input. Previous studies have suggested that attention affects speech processing in infants (see Jespersen 1997). The attainment of native-like speech perception between 9 and 11 months (as reflected in decreased sensitivity to non-native contrasts to which infants have had no exposure) has been linked to performance on problem-solving tasks that require attentional control (Conboy, Sommerville, & Kuhl 2006; Lalonde & Werker 1995). This suggests an underlying role for attention in both sets of skills. Perhaps infants’ attention to audiovisual speaker cues, shown in previous studies to be important for phonetic learning (Kuhl & Meltzoff 1982, 1992), was enhanced during the live interactions but not during passive viewing of the DVDs. Indeed, attention rating scores on a 4-point scale indicated that infants in the live condition were more attentive to the speakers and their materials than those in the audio–visual group, and the latter group was more attentive than infants in the audio-only condition. Kuhl and colleagues have argued that attention and motivation are key elements for communicative learning in humans and other species, and are enhanced by social contact (Doupe & Kuhl 1999; Kuhl et al. 2003, 2007).

For example, several studies of songbirds have demonstrated the importance of live interaction from tutors for learning (Adret 1993; Baptista & Petrinovich 1986; Eales 1989; Immelmann 1969; Tchernichovski, Mitra, Lints, & Nottebohm 2001).

The social and cultural relevance of shared speech perception

According to Rogoff (2003), human development takes place through participation in cultural communities, and can only be understood in cultural terms. We propose that one aspect of human development, phonetic learning, is shaped through a sharing of perception in social and cultural contexts. The embedding of phonetic information in meaningful communicative interactions motivates learning in infants by enhancing attention to relevant acoustic features. Social cues are generated during dynamic, live interactions found in the successful instance of second language learning from live short-term exposure reported by Kuhl et al. (2003), and are also present in first language acquisition contexts. The 12 25-minute sessions used in the Kuhl et al. (2003) study were not completely naturalistic – the speakers talking to the infants followed a script while reading books and played with preselected toys – but they closely simulated learning in the real world, and were very distinct from studies of statistical learning in which infants listen to synthetic syllables presented auditorily for short periods of time. Thus the social cues generated when these sessions were live, as opposed to pre-recorded, may have been necessary for successful intake of the complex audiovisual information provided. Social cues are important for language learning throughout the first year, but the importance of particular cues may increase with age as advances in social cognition allow infants to make better use of such information and as learning environments become more complex.

The role of social interaction in infant language acquisition has been discussed by numerous scholars (e.g., Bates 1976; Bloom 1993; Bornstein 1996; Bruner 1983; Gallaway & Richards 1994; Hart & Risley 1995, 1999; Roodenreis 1998; Snow 1977, 1999; Snow & Ferguson 1977; Tomasello 2003; Trevarthen 1998; Vygotsky 1978). A growing body of evidence has suggested that face-to-face communicative interactions occurring between infants and their caregivers in the first months of life set the stage for subsequent social, cognitive, and language development (e.g., Baumwell, Tamis-LeMonda & Bornstein 1997; Beckwith et al. 1976; Beckwith & Rodning 1996; Bornstein et al. 1990; Bornstein & Tamis-LeMonda 1989, 1997, 2001; Britten 1998b; Clarke-Stewart 1973; Jaffe et al. 2001; Klein, Weider & Greenspan 1987; Nicely, Tamis-LeMonda, & Bornstein 1999; Sigman & Beckwith 1980; Tamis-LeMonda & Bornstein 2002; Tamis-LeMonda, Bornstein, & Baumwell 2001; Trevarthen & Aitkin 2001). During the earliest face-to-face interactions infants and adults mutually and reciprocally attend and attune to each
other's states with remarkably high levels of coordination (Batson 1979; Beebe et al. 1988; Crown et al. 2002; Iaffo, Stern, & Peery 1973; Iaffo et al. 2001; Murray & Trevarthen 1986; Stern 1977; Stern et al. 1975; Trevarthen 1979, 1998). The spontaneous, rhythmic "protoconversations" that take place during the first few months of life, marked by mutual gaze, smiling, and vocalizations (Batson 1979; Stern 2002), appear to reflect the infant's desire for communication with other humans, and thus prepare infants for the acquisition of language (Trevarthen 1998). Moreover, they are meaningful in the sense that they are embedded in a collective, a community of meaning provided by the adult (Rommetveit 1998). Thus, even during the earliest stages of infancy, a primary intersubjectivity entailing some sense of a "virtual other" is present, and appears to form a foundation for learning via participation with other humans (Braten 1998b).

Such reciprocal attunement with other humans may underlie the infant's earliest successes at learning to perceive speech in culturally specific ways. As early as 6 months, the perception of vowels reflects the influence of the language environment (Kuhl et al. 1992). In a previous chapter, Kuhl (1998) suggested that an ability to share perception with other speakers of their language allows such early language-specific learning to take place. In that chapter, it was argued that a sharing of perception allows infants to "mentally align themselves with adults of the culture" (p. 297). In other words, infants learn mental maps, or filters, through which to perceive language, which resemble the mental maps of other members of the language community. Exposure to the sounds used by the community's language and distributional properties of those sounds in the ambient language provide additional cues that allow infants to form such culturally specific perceptual filters or "native language perceptual magnets" (Kuhl 1993, 2000). Thus when the infant experiences multiple exemplars of a phonetic prototype (the most representative instance of a phonetic category), his/her perception of other speech sounds in nearby acoustic space is influenced by that prototype. The prototype functions as a "perceptual magnet," perceptually pulling the other sounds towards it so that all of the sounds are perceived as members of the same category by speakers of that language. The infant becomes a member of the language community by sharing in this language-specific perceptual warping.

At the time of that writing, little was known about the nature of the language experience needed for such learning to occur. It had been established that infants as young as 6-8 months of age could learn certain properties of a language from passive exposure to the statistical properties present in disembodied speech directed at them during structured experiments (e.g., Aslin et al. 1998; Goodlett et al. 1993; Maye et al. 2002; Saffran et al. 1996, 1997). It had also been established that a yoking of visual information (mouth shape) to the auditory information in speech was important for perception (Kuhl & Meltzoff 1982, 1996). The more recent finding that 9-10-month-olds do not learn phonetically from passive exposure to more naturalistic forms of speech presented in audiovisual formats (Kuhl et al. 2003) compel us to address (a) the differences between passive exposure to a language and experience with language in meaningful sociocultural contexts, and (b) the different ways language learning occurs at different developmental stages. We have already discussed what it means for an infant to experience language in natural communicative interactions with members of their language community. Such protoconversations are characterized by reciprocal attunement and attention to each other's states. We now turn to what it means to experience language towards the end of the first year of infancy, when advances in cognitive skills allow infants to take in more information and participate more fully in a sharing of perception.

Social-cognitive factors in the development of speech perception

Shifts in social cognition occurring in the second half of the first year have been linked to important transitions in language development. Most notably, the infant's increasing ability to understand another person's reference to an object of joint attention is crucial for the acquisition of a meaningful, referential lexicon (Akhtar & Tomasello 1998; Bakeman & Adamson 1984; Baldwin 1995; Baldwin & Markman 1989; Brooks & Meltzoff 2002; Bruner 1983; Carpenter, Nagell, & Tomasello 1998; Gogate, Walker-Andrews & Bahrick 2001; Tomasello 1999; Tomasello & Farrar 1986; Tomasello & Todd 1983). Several key developments coincide with this ability to understand reference. By 9 months infants begin to engage in triadic "person-person-object games" – they systematically combine purposes directed to objects with those that invoke interest from another human, reflecting a "secondary intersubjectivity" (Trevarthen & Hubley 1978; Trevarthen 1998). Braten (1998a) has demonstrated that towards the end of the first year, infants display an "altercentric" participatory perception of others, as seen in behaviours that require a reversal of the perspective of the other during face-to-face interactions. Tomasello and colleagues have further argued that shared perception of communicative intentions, which emerges at around 9 months of age, is crucial for the acquisition of language (Akhtar & Tomasello 1998; Tomasello 1999). The ability to attend to objects of another person's reference appears to be linked to the infant's ability to understand others as intentional agents (Tomasello 1999). Around the same time that infants begin to display such abilities, they also begin to display comprehension of the meanings of words (Fenson et al. 1994).

We suggest 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. For example, 9-10-month-old infants can follow the line of regard of others. When
faced with a language learning situation, specific meaningful social cues provided by adults, such as eye gaze and pointing to an object of reference, may help infants segment words from ongoing speech, thus facilitating phonetic learning from the sounds contained in those words (Kuhl et al. 2003). This does not necessarily mean that infants this age are able to integrate all of the relevant information present in the signal at once. Indeed, recent studies have suggested that even at 14 months, infants are unable to use fine phonetic detail when processing words with meaning (Mills et al. 2004; Stager & Werker 1997; Werker, Fennell, Corcoran, & Stager 2002), although infants this age can perceive such phonetic information in words (Swingley & Aslin 2002). Early language acquisition is most likely a piecemeal process in which multiple pieces of information are gradually integrated (Hollich, Hirsh-Pasek, & Golinkoff 2000). When faced with a new language learning situation, infants this age may tune into the meaningfulness of interactions with their interlocutors first, and this in turn may facilitate their learning of other aspects of the language. Situations in which meaning is mediated by the adult, via behaviours that are contingent on the infants' behaviours, are thus more likely to induce learning than situations not mediated in this way. Infants' understanding of the contingencies between their actions and those of adults may influence their attention to linguistic units at all levels. Infants' behaviours, such as levels of general attention and arousal, may in turn enhance the quality of the input they receive from adults.

The role of shared perception for phonetic learning may increase throughout development, as infants become increasingly aware of the communicative intentions of others and more in tune with the meaning of such communication. As such, older infants may be more attuned to information presented in situations in which their interlocutors' responses are contingent upon their actions. Further research is needed to determine whether audiovisual information presented in a static, non-contingent format such as the DVD condition in the study reported by Kuhl et al. 2003) is more useful to infants at particular developmental stages, and whether a television format in which such contingencies are present would be equally effective for phonetic learning. In a study of 14- and 24-month-old infants, Meltzoff found that infants of both ages could imitate adult actions with objects presented over a television screen, even after a 24-hour delay (Meltzoff 1988). In that study the video presentations were timed so that they coincided with optimal levels of infant visual attention. Language exposure presented in a similar format, while not completely interactive, could nevertheless provide some level of responsivity to infants' attentional levels.

The phonetic learning that occurs in live, natural interactions with speakers of a language may also be more robust and durable than that which occurs through static, non-contingent exposure to speech. For example, infants in the live Mandarin exposure experiment were tested between 2 and 12 days after the final exposure session. A median split based on the number of days between infants' last exposure session and the HT learning assessment session indicated that learning was not short-lived. Infants tested immediately after the last exposure performed similarly to those who were tested more than a week later. Learning in live interactions and controlled experiments may be fruitfully compared with regard to the durability and robustness of learning. Our working hypothesis is that infants need exposure to multiple instances spoken by different talkers in natural settings to show phonetic learning that is robust and durable. Even the 5-hour exposure that infants experience in our experiments is not expected to be as robust as the learning shown by infants who have been raised listening to a particular language for 11 months.

Further research is also needed to determine whether there are individual differences in the ability to shift to culturally/linguistically appropriate ways of listening to a language. As discussed previously, studies from our lab have shown that individual variation in the attainment of native language perception is linked to other advances (Conboy et al. 2005; Kuhl et al. 2005; Rivera-Gaxiola et al. 2005b; Tao et al. 2004). Individual variation in the learning of a new language may also be predicted by first language abilities, and by advances in other areas of cognition.

For example, infants who display more joint attention behaviours during language learning situations may learn more than those who hear the same phonetic information but engage in fewer joint attention behaviours. Infants who display a greater number of contingent responses to the adults' actions and speech may show a greater amount of learning than other infants. Furthermore, infants who are more advanced on problem-solving tasks that require high levels of attentional control may be better able to learn from second language exposure than infants who are less advanced on such tasks.

Current research and future directions

We are currently exploring these ideas through a study in which infants from English-speaking homes are being exposed to Spanish at 9–10 months of age. Infants are brought to the lab for a total of twelve 25-minute sessions in which they hear Spanish from native speakers who interact with them naturally while showing them toys and picture books. In this research, we are also analyzing the social behaviours that occur in these second-language learning situations. Of interest to us is whether aspects of the interactions are predictive of the amounts of phonetic learning that occur. We are assessing the robustness and durability of learning by comparing performance on the HT task between infants who receive this short-term exposure to Spanish and those who have had naturalistic exposure to Spanish for longer periods of time, both in monolingual and bilingual learning situations.
We are also exploring whether infants are able to learn other aspects of language from these interactions. If they are attending to their language input at the word level, then this should also be reflected in the way they process words. Previous research using event-related brain potentials (ERPs) has established that infants as young as 9–11 months of age process words they have been familiarized with differently from unfamiliar words (Addy & Mills 2005; Mills, Coffey-Corina, & Neville 1994; Thierry, Vihman, & Roberts 2003). If infants attend to language at this level in second-language learning situations, then the ERPs to words they have heard during these sessions should be processed differently from words they have never heard. As with phonetic learning, such lexically based learning may be predicted by the social behaviours displayed by infants and their adult communicative partners during second-language learning interactions.

We are also examining whether individual differences in the overall quality of dyadic behaviours during language exposure sessions are linked to later cognitive and language outcomes. Previous research has shown that infants who receive high levels of "mediated learning experiences" from their mothers show better performance on tests of language and cognitive development at 2, 3, 4, and 5 years of age (Klein & Alony 1992; Klein, Weider, & Greenspan 1987). In this research, interactions were said to be "mediated" by the adult when they contained the following elements: (1) intentionality (an act directed toward affecting the infant's perception or behaviour); (2) mediating of meaning (expressions of excitement, appreciation, or affect in relation to objects, concepts, etc.); (3) transcendence (attempts to expand the infant's cognitive awareness); (4) mediation of feelings of competence (expressions of satisfaction with the infant's behaviour); and (5) regulation of behaviour (matching tasks to the infant's capacities, interests, etc.). Such experiences may be defined differently across cultures, but are believed to be found universally. Other research has shown that adult responsiveness to infants predicts cognitive and language outcomes (e.g., Baumwell et al. 1997; Beckwith et al. 1997; Beckwith & Cohen 1989; Bloom 1993; Bornstein 1989; Bornstein et al. 1990; Bornstein & Tamis-LeMonda 1989, 1997; Bornstein, Tamis-LeMonda, & Haynes 1999; Carew & Clarke-Stewart 1980; Clarke-Stewart 1973; Landry et al. 1997, 2001; Nicely, Tamis-LeMonda & Bornstein 1999; Tamis-LeMonda & Bornstein 2002; Tamis-LeMonda et al. 1996, 2001). Based on this previous work and the results of our studies reviewed in this chapter, we predict that both the quality of infant-tutor interactions during language exposure, as measured by the amount of mediation provided by the adult and the responsiveness of the adult to the infant's behaviours, will influence infant second language learning. On the infant's part, joint attention and engagement behaviours should enhance learning.

A final note about cultural diversity in language learning

We have suggested that social interactions between infants and adults provide a cultural context that moves infants into specific ways of listening to language and facilitates subsequent language learning from members of the cultural linguistic group. A key component of these interactions may be the degree to which infants and adults engage in contingent behaviours that enhance infants' attention to relevant linguistic information. Thus, infants' cognitive abilities, their understanding of social cues, and adult behaviors coincide to provide effective environments for phonetic learning. We have presented data that suggest how these factors might influence second language phonetic learning in infancy, and have generated hypotheses we are now investigating. We have also suggested that similar processes play an important role in first language phonetic learning. Additional research is needed to examine what kinds of phonetic information can be learned across first and early second language situations, and under what conditions. More work is also needed to assess the extent to which short-term exposure to language under conditions that simulate natural language interactions can result in robust, durable learning such as that found in primary language acquisition.

We also wish to emphasize that there are likely to be differences across cultural groups in the ways in which phonetic learning occurs. Face-to-face interactions between infants and adults are not uniformly common across all cultures (e.g., Bornstein et al. 1990a, b, c; Heath 1983; Martini & Kirkpatrick 1981; Rogoff 2003; Schiefelbein 1991); thus, they may not be as important for the development of shared perception as has been suggested by development research in middle-class Western cultures. The degree to which talk is valued compared to nonverbal forms of communication varies across cultures. Distal forms of communication involving sound, rather than more proximal nonverbal communication involving touch, may be emphasized to a greater extent in cultures in which infants are physically separated from other people compared to those in which infants are always kept close (see Rogoff 2003). It is well known that across many cultures the speech directed towards infants (often referred to as "motherese") contains properties that attract and hold infant attention (Fernald 1984; Fernald & Kuhl 1989), and speech units that are exaggerated (Kuhl et al. 1997). Our work has recently shown that the degree of mothers' speech clarity in these infant-directed episodes strongly correlates with infants' speech discrimination in the HT task (Liu, Kuhl, & Tsao 2003), suggesting that the exaggerated language directed towards infants in social settings attracts infants' attention and assists learning. Whether this process is consistent or different across cultures remains to be explored. In some cultures infants are not regarded as conversational partners, and little speech is directed at them (Heath 1983; Ochs & Schieffelin 1984; Schieffelin 1991; Ward 1971). However, the infants are surrounded by the speech of others, and their actions or interests
may be commented on by others in a contingent fashion (Schieffelin 1991). Infants in such communities may learn quite a lot about language by "listening in" or "eavesdropping" (Rogoff 2003). Older siblings and other children also function as important agents of language socialization in some cultures (Schieffelin 1991; Ward 1971; Zukow 1989); overlooking their influence could result in a misunderstanding of the developmental process that goes toward the development of speech perception. Dyadic interactions are also rare in some cultures (Rogoff et al. 1993; Whaley et al. 2002); in such cases, analyses at the level of the dyad would not be appropriate. Levels and types of participatory learning tend to vary across members of any given culture, and must be viewed as a cohesive whole to be adequately understood (Rogoff 2003; Whaley et al. 2002). Thus when we look at the ways in which speech perception is shared amongst members of a community, we must consider the interactions that occur at many levels within the community. Future research should consider the social processes that occur amongst members of diverse cultural communities in order to understand the underlying mechanisms underlying the development of native language speech perception.

References


