2

Evolution, Nativism and Learning in the Development of Language and Speech

Patricia K. Kuhl and Andrew N. Meltzoff

Introduction

Infants acquire language like clockwork. Whether a baby is born in Stockholm, Tokyo, Zimbabwe or Seattle, at 3 months of age, a typically developing infant will coo. At about 7 months the baby will babble. By their first birthday, infants will have produced their first words, and by 18 months, 2-word combinations. Children of all cultures know enough about language to carry on an intricate conversation by 3 years of age.

When our own daughter began to produce the "babababa" characteristic of canonical babbling, we were struck by the regularity of its form and the precision of its timing. Having occurred on schedule rather than being accelerated by our ever-constant modelling, we were reminded that the milestones of human language occur at the appointed time regardless of the language in which the child is being reared, the educational background of the infant's parents – and, apparently, regardless of parental prompting or the theories they hold.

Such observations seem to support Chomsky's nativist view that language milestones occur at pre-specified times, as do the eruption of teeth or the onset of puberty. Recent discoveries, however, require revisions to this idea. The emerging view remains strongly nativist, to be sure, but suggests a critical role for language input. The new view provides some insight into how one *particular* language rather than another is acquired. Not only the fact that infants are language-generalists needs explanation (Chomsky's forte), but also the process by which they so quickly become culture-bound language-specialists, adopting a particular "native tongue" that permanently marks them. This indelible mark presents one of the deepest mysteries of early language development: try as one might, unlearning the accent or phonology of one's native tongue is virtually impossible. Henry Kissinger was not born with a German accent, nor Chomsky born with a Philadelphian one. These are not innate characteristics; once acquired, however, they have persisted over decades. Such is the mark of early learning.

The new data also suggest another shift from the standard nativist view. During early development, there is no compelling reason to postulate that the linguistic system functions independently of other cognitive and social systems. We will argue that although the language system may become modularized with development, infants do not begin life with a fully organized language module that is isolated from other aspects of cognition (Fodor 1983).

We will suggest a view that incorporates evolution, nativism and experience in the development of language. Our view embraces the notion that infants are born with abilities highly conducive to the development of language. We are nativists in this sense. These innate abilities *initially structure* the acquisition of language. However, infants' innate abilities do not *solely determine* language. Linguistic experience alters the system in profound ways. It fully restructures the system, and does so quickly, relatively permanently, and via an interesting mechanism that will be described here.

The theory and the arguments we present primarily address the phonetic level of language, the perception and production of the most basic units of language, the consonants and vowels of human speech. The phonetic level has advantages: one can study the comparative, developmental and cross-cultural aspects of the perception and production of speech. Even machines' capabilities to categorize the sounds of language can be tested. It therefore allows a comprehensive look at the underpinnings of humans' linguistic capacity. Our hope is that study of the phonetic level of language may inform theories of language acquisition at other levels.

The more specific goal is to elaborate further the *Native Language Magnet* theory of speech development first described by Kuhl (1992a, 1992b, 1993a, 1993b, 1994). A three-step process in the acquisition of speech is postulated: (a) innate perceptual boundaries exist that are tailor-made for language processing at the phonetic level; (b) exposure to ambient language results in stored representations that reflect the distributional properties of a particular language; and (c) the stored representations act recursively to alter the innately specified boundaries; they profoundly influence the subsequent perception and production of speech in relatively permanent ways. We believe that in early infancy, language acquisition is underpinned by a more general cognitive representational ability like the one described by Meltzoff (1990). This early representational system is polymodal – it is one to which all sensory modalities as well as the motor system has access. Moreover, the type of experience that influences speech representation entails a rather special interaction that occurs with conspecifics (Meltzoff and Gopnik 1993; Meltzoff and Moore 1995); a tape recorder presenting the

sounds of language would not trigger it. The specific interweaving of what is "given by nature" and what is "gained from experience" is the story we will tell.

Nativism: Initial Structure for Phonetic Categorization

Infants have innate perceptual abilities that support the acquisition of language at the level of speech. Two pieces of evidence stand out: (a) categorical perception, a phenomenon showing that infants' perceptual systems partition sound to roughly define the phonetic categories of language; and (b) talker normalization, a phenomenon demonstrating that infants perceive their own vocalizations as "matching" adults' vocalizations, even though the two are physically very different. Even the most sophisticated computers have not succeeded in this special capacity for talker normalization. Yet human infants do so with ease. Such a biological endowment is necessary for infants to acquire the ability to speak themselves.

Categorical Perception

Tests of categorical perception (CP) use a continuum of speech sounds as stimuli. A series of sounds is generated by altering some acoustic variable in small steps. On one end of the series the sounds are identified as one syllable, the syllable /ba/ for example; on the other end of the continuum the sounds are identified as another sound, the syllable /pa/ (Liberman, Cooper, Shankweiler and Studdert-Kennedy 1967) (Figure 1).

Tests of CP ask listeners to identify each one of the sounds in the series. Early researchers expected that the sounds in the series would be perceived as changing gradually from /ba/ to /pa/, with many sounds in the middle of the series sounding "ambiguous" or a poor mixture of the two. That did not occur. Adults reported hearing a series of /ba/'s that abruptly changed to a series of /pa/'s. There was no in-between. When researchers asked listeners if they could hear the difference between two adjacent /ba/'s (or /pa/'s) in the series, they could *not* do so, even though the two /ba/'s (or /pa/'s) were physically different. Listeners did not hear differences between adjacent stimuli in the series until they heard a sudden shift – the change from /ba/ to /pa/. The fact that listeners' responses were "categorical" gave the phenomenon its name.

CP is sensitive to linguistic experience (Miyawaki, Strange, Verbrugge, Liberman, Jenkins and Fujimura 1975). For adults, CP occurs only for sounds in their native language. When Japanese listeners were tested on a series of sounds that ranged from /ra/ to /la/ (a distinction that is not phonemic in Japanese), they did not hear a sudden change

An acoustic continuum
with equal physical steps

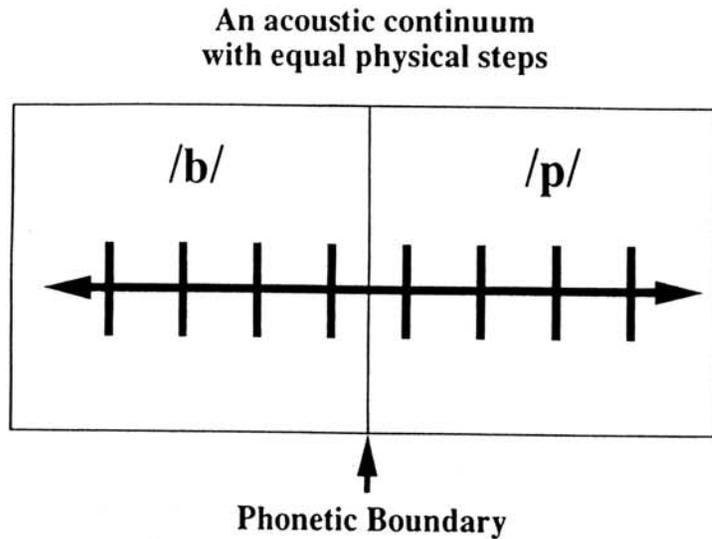/b/   /p/

Phonetic Boundary

Figure 1: Categorical perception is tested using sounds from a computer-generated series. The sounds vary in equal steps along an acoustic dimension; however, perception changes abruptly at the location of the phonetic boundary between the two categories.

at the boundary between /ra/ and /la/. They heard no change at all. (This is why Japanese speakers tend to substitute /l/ for /r/ in speech.) Nonetheless, American listeners reported hearing a series of /ra/'s that changed suddenly to a series of /la/'s (Figure 2, top). The bottom half of Figure 2 compares the American and Japanese discrimination data. American listeners showed the characteristic peak in discrimination at the location of the /r-l/ boundary; Japanese listeners did not show this peak in discrimination at the phonetic boundary. Their performance in discriminating /ra/ from /la/ was at chance throughout the series (Figure 2, bottom).

The Americans were unlikely to have one set of innate endowments (a /ra/-/la/ detector) and Japanese another; that CP was language-specific suggested that it might be learned. Perhaps this learning arose as a result of hearing words with different referents contrasting /b/ and /p/ – like "bat" and "pat." If so, then very young infants would not be expected to show CP.

The relevant study was done by Eimas, Siqueland, Jusczyk and Vigorito (1971). Infants' responses to a /ba-pa/ series were monitored using a specially designed technique that relied on the measurement of sucking. The results showed that young infants demonstrated CP. Moreover, infants demonstrated the phenomenon not only for the
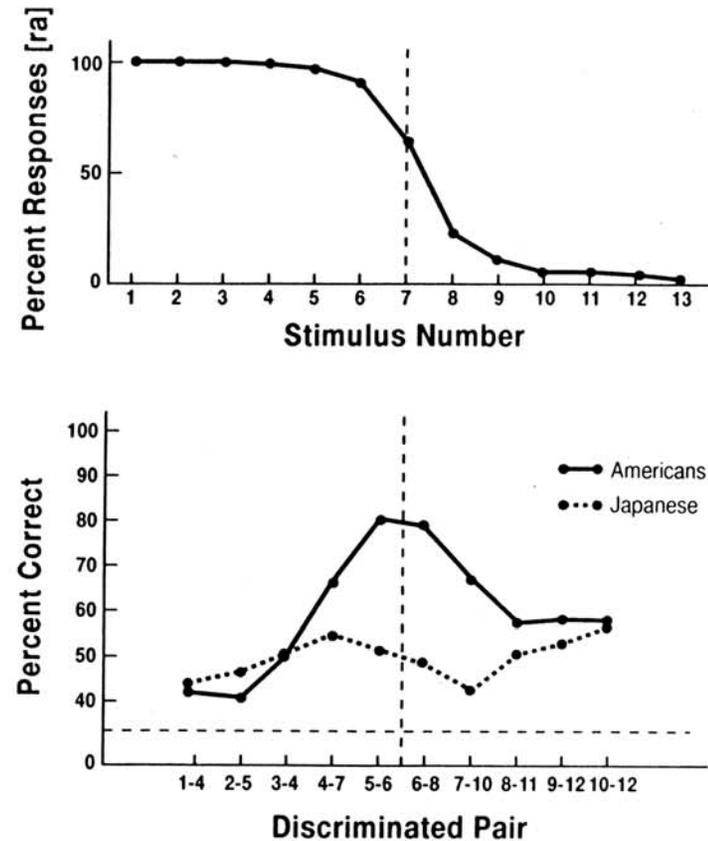
Figure 2: An example of categorical perception for the syllables /ra/ and /la/. American listeners identify sounds in the series (top) and show the characteristic peak in discrimination at the location of the phonetic boundary (bottom). Japanese listeners do not show the peak in discrimination. Redrawn, with permission, from K. Miyawaki, W. Strange, R. Verbrugge, A.M. Liberman, J.J. Jenkins and O. Fujimura (1975), An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics* 18: 331–40.

sounds of their own native language, but also for sounds from many foreign languages (Streeter 1976, Lasky, Syrdal-Lasky and Klein 1975). Although adults were "culture bound," infants were not: they were primed to be members of any linguistic culture, "citizens of the world."

It can be concluded that infants' auditory perception is tailored to language processing at birth. Moreover, this does not depend on

experience. Infants behave this way even for sounds they have never before heard. The puzzle that remains (to be discussed later) is when, why and by what mechanism adults "lose" a language-related ability that is present at birth.

Talker Normalization

The studies on CP show some rudimentary structure available to infants that helps them partition the perceptual space into gross divisions. However, perception of a phonetic category requires something more. In order to perceive a phonetic category, infants have to be able to perceive *similarity* among sounds that belong to a particular category, even though they are discriminably different. When different people produce the same vowel sound, one can hear the differences between them but one can also hear their identity. This is phonetic constancy despite auditory discriminability, categorization that renders discriminably different things equivalent.

This categorization ability is critical to infants' acquisition of speech. Infants' vocal tracts cannot produce the frequencies produced by an adult's vocal tract, so they cannot create the exact frequencies that an adult produces. Infants must hear the commonality between the vowels they are capable of producing and those produced by adults in order to learn to speak. Computers cannot yet be programmed to "perceive" these kinds of similarities across a wide range of talkers. Would naïve infants outperform the smartest computers in perceiving a perceptual similarity, *constancy*, for the same vowel produced by different talkers?

Kuhl demonstrated that infants have the ability to sort vowels by phonetic category regardless of the talker producing the sound (Kuhl 1979; 1991b). Figure 3 shows the results of two studies. Shown in the top panel are infant data from an /a-i/ categorization experiment (Kuhl 1979), and in the bottom panel, results from an /a-ae/ categorization experiment (Kuhl 1991b). In both, infants were initially trained to produce a head turn to a single vowel from Category 1 produced by a male speaker, but not to produce the head turn to a single vowel from Category 2 by the same speaker. The first and third panels show the results of the training data; infants master this task at the 90% correct level in short order. During the test phase of the experiment, novel exemplars are presented from both Category 1 and 2, produced by new male, female and child talkers. The results of these studies (second and fourth panels) demonstrated that infants generalize their head-turn response to the novel vowels of Category 1, but not Category 2, which is predicted by the hypothesis that infants are capable of perceptually sorting the novel vowels into two phonetic categories.
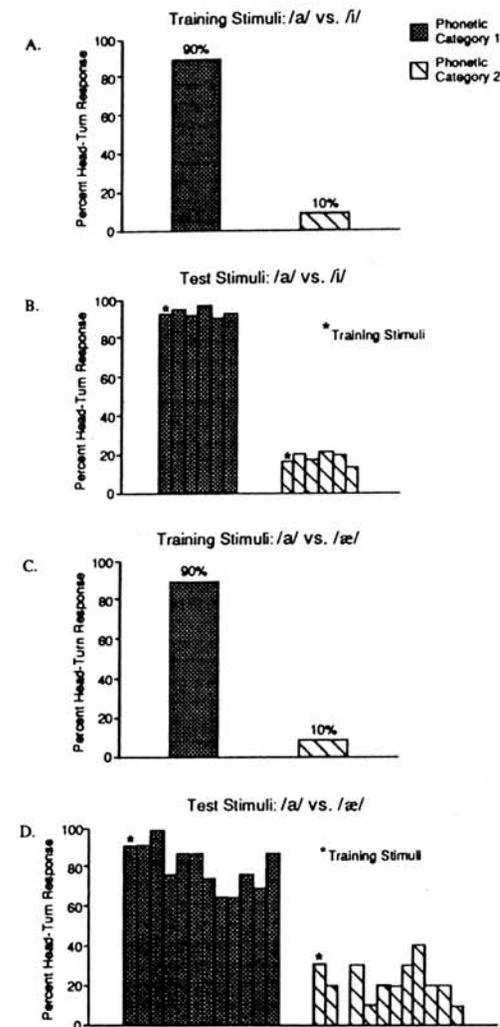

Figure 3: Categorization data from 6-month-old infants. Infants are trained until they reach 90% correct performance on the discrimination of two vowels spoken by a single speaker, panel A (/a/ vs. /i/), panel C (/a/ vs. /ae/). Infants were then tested using vowels produced by many speakers, including new male, female, and child speakers (panels B and D). Performance indicates that infants can perceptually sort vowels into phonetic categories regardless of the speaker who produces the vowel. From P.K. Kuhl (1991b), Perception, cognition, and the ontogenetic and phylogenetic emergence of human speech. In S.E. Brauth, W.S. Hall and R.J. Dooling (eds.), *Plasticity of Development*, 73–106. Cambridge, MA: MIT Press.

Infants succeeded for both relatively easy vowel contrasts such as /a/ versus /i/, and difficult contrasts such as /a/ (as in "pot") versus /ae/ (as in "pat"). In the /a -ae/ case, the vowels were naturally produced by 12 different men, women and children. Voices that sounded very different were purposely chosen. Women with exceptionally high voices, men with deep voices, even people with colds who sounded very nasal but could be understood. The stimuli would have confused even the most complex computer designed to categorize phonetic units. Infants had no trouble sorting these vowels into categories despite these acoustic differences. Talker normalization has now been shown in 2-month-old infants (Marean, Werner and Kuhl 1992) and newborns (Walton, Shoup-Pecenka and Bower 1991).

These two perceptual abilities – CP and perceptual constancy – are innate foundations for speech and language learning. The CP phenomenon shows that infants parse the sound stream in a way that segments the basic units of speech. CP provides "basic cuts" in the acoustic stream that coincide with linguistic categories. Talker normalization provides another benchmark. The physical (acoustic) disparity between the voices of different individuals is so extreme that it prevents computers from correctly categorizing speech across a wide range of talkers; it is still unknown what makes for "/i/-ness" in a vowel sound spoken by different talkers – the essence of /i/ cannot be identified by any known algorithm. But the biological mechanism available to infants picks out the /i/-ness of a vowel despite who says it. It recognizes phonetic units that remain invariant or constant despite the acoustic differences of gender, colds and the like. The combined findings of CP and vowel categorization constitute strong evidence that infants are evolutionarily prepared for language acquisition.

Effects of Experience: Stored Representations and Formation of the Brain's Perceptual Maps

Infants are innately prepared to hear the sounds of a universal language. Adults' perception of speech is more restricted and culture-bound, which suggests an odd or reverse sort of learning. One of us, P.K. Kuhl, was graphically reminded of this. She was visiting a speech laboratory in Japan, preparing to test Japanese infants' perception of the American English /r/ and /l/ sounds. As the stimuli began to play out of the loudspeaker, her seven Japanese colleagues (1 professor, 3 graduate students and 3 undergraduate students) gathered in the sound-proof booth. Kuhl listened as crystal clear versions of /ra/ and /la/ played from the loudspeaker, pleased that the computer disk had survived the trip and that the experiment was ready to run. She looked at her Japanese colleagues as they quizzically looked at each other.

"Which is it?" one finally queried. Not a single one of the Japanese adults could identify whether the sound coming out of the loudspeaker was /ra/ or /la/, nor even identify when the sound changed from one to the other. This was true even though all of them understood a certain amount of English, and could communicate with Kuhl. It was a powerful reminder that the effects of language experience leave a mark on our perceptual abilities.

A similar example is that of American English listeners who have great difficulty hearing the difference between the Spanish /b/ and /p/, sounds that are perceived as belonging to the same phonetic category (/b/) in American English but are easily distinguished by Spanish listeners (Abramson and Lisker 1970). These examples show how the "language-general" pattern of phonetic perception we possessed as infants has become "language-specific." When, how and why does this happen?

Werker and her colleagues showed that by the end of the first year of life there is a change in infants' perception of foreign-language phonetic contrasts (Werker and Tees 1984; Werker and Lalonde 1988; Werker and Pegg 1992). At this age, infants demonstrated a failure to discriminate foreign contrasts that they earlier showed an ability to discriminate. It was suggested that it might be mediated by the acquisition of word meaning (Werker 1991). It was thought that by 12 months, infants had begun to learn which sounds made a difference in their language and that they had began to ignore the phonetic variations that did not make a difference in word meaning.

However, more recent results from Kuhl's laboratory show that infants' perception of speech is altered by language exposure much earlier in life, which radically altered our view of the mechanism underlying this change. The new findings show that by 6 months, exposure to language has already altered infants' perception of speech (Kuhl, Williams, Lacerda, Stevens and Lindblom 1992). This new finding suggests that the change in infants' perception of speech does not depend on the acquisition of word meaning. What is the nature of this change and how is it brought about?

Phonetic Prototypes

Recent work in Kuhl's laboratory has produced an effect that helps explain how language experience alters speech perception and production. The effect shows that language experience alters the perceived distances between speech stimuli – that it, in effect, "warps" the perceptual space underlying speech.

The effect, termed the *perceptual magnet effect*, was uncovered in experiments using phonetic "prototypes," the best or most representative

instances of a phonetic category (Kuhl 1991a; Kuhl et al. 1992; Kuhl 1993a, 1993b, 1993c, 1994; Iverson and Kuhl 1995, in press). Experiments on visual prototypes were orignially done by Rosch, who defined them as "good instances" of categories, instances that are representative of the category as a whole (Rosch 1975, 1978; Posner and Keele 1968). It has been demonstrated that the prototypes of categories are special – they are easier to classify, easier to remember, and often preferred over other members of a category (Mervis and Rosch 1981; Rosch 1975, 1977).

Initial studies in Kuhl's laboratory on phonetic prototypes were undertaken with adults to establish whether listeners perceived that the members of speech categories differed in quality (Kuhl 1991a). Three discoveries were made: (a) listeners are very good at identifying phonetic prototypes, sounds that were the best instances of the category; (b) phonetic prototypes are language specific in adults; and (c) phonetic prototypes have a unique function in perception, acting as "magnets" for other sounds in the category.

First, studies revealed that adult listeners were very good at identifying best instances or prototypes of the consonants and vowels of their native language (Davis and Kuhl 1992, 1993; Grieser and Kuhl 1989; Iverson and Kuhl 1995, in press; Kuhl 1991a, 1992a). Listeners' goodness ratings revealed "hot spots," places in acoustic space where ratings for a particular category were very high. As one moved away from that location, the ratings consistently dropped. Moreover, these ratings were language specific. American listeners had hot spots in places the Swedes did not and vice versa (Kuhl 1992b). This was true even for the same phonetic unit. For example, the /i/ judged best by the Swedes was located in a different place than the /i/ judged best by the Americans. The data suggested that the adults of different languages mapped the vowel space very differently, with varying numbers and locations of vowel hot spots.

A second finding revealed the psychological effect of the prototype. Prototypes acted as "perceptual magnets" for other sounds in the phonetic category. When listeners heard a prototype of a phonetic category and were asked to compare the prototype to similar sounds that surrounded it in an acoustic space (Figure 4 A), the prototype displayed an attractor effect on the sounds (Figure 4 B) (Kuhl 1991a). The prototype perceptually pulled other members of the category towards it. Poor instances from the same category (non-prototypes) did not function in this way.

How does the magnet effect work? Studies using multidimensional scaling (MDS) techniques reveal that the magnet effect distorts the perceptual space underlying a phonetic category (Iverson and Kuhl 1995, in press; Kuhl and Iverson 1995). Specifically, MDS was used
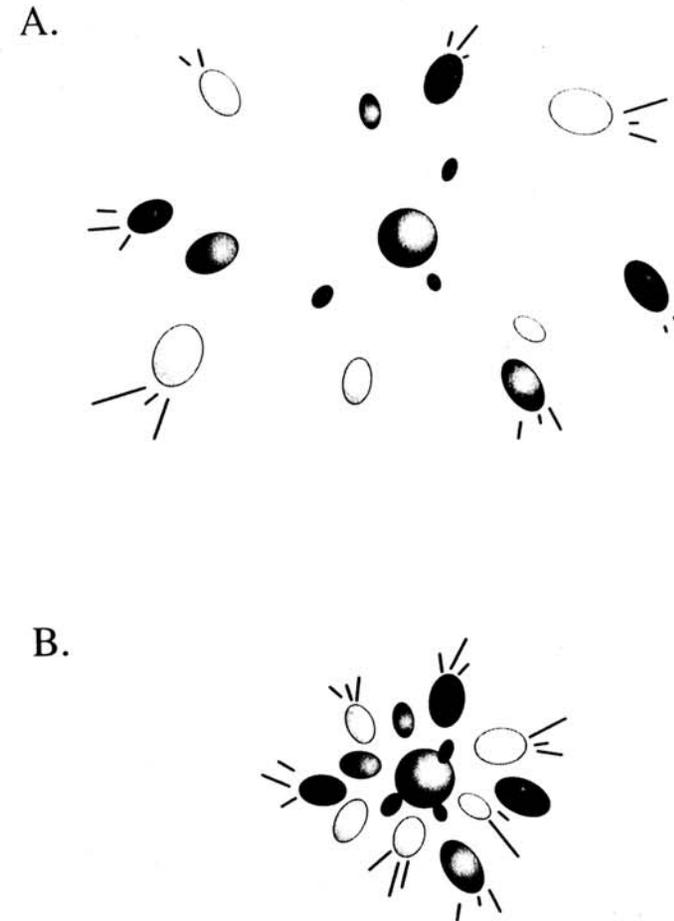

A.

B.

Figure 4: The perceptual magnet effect: When a variety of sounds in a category surround the category prototype (A), they are perceptually drawn towards the prototype. The prototype appears to function like a magnet for other stimuli in the category. From P.K. Kuhl (1993b), Infant speech perception: A window on psycholinguistic development. *International Journal of Psycholinguistics* 9: 33–56.

to assess potential *contraction* and *expansion* of the perceptual space underlying vowels.

Subjects were tested with vowel stimuli spaced at equal physical distances in vowel space. Listeners first identified the stimuli as either the /i/ in the word "he" or /e/ in the word "hey." They then rated the category goodness of each vowel on a scale from 1 (poor) to 7 (excellent). Finally, they listened to all pairs of the 13 stimuli and judged whether the tokens in each pair were the same or different. If a pair of sounds were very different, the subject's reaction time (RT) was quite short; if the pair of sounds was very similar, the RT was relatively long. Subjects' responses were analyzed using multidimensional scaling techniques (Kruskal 1964) which organize RTs in a spatial array so that pairs of stimuli with long RTs (high similarity) are placed close together, while tokens with short RTs (low similarity) are placed far apart.

The magnet effect predicts a tight clustering in space in the region of the best instances (prototypes), and separation in space for stimuli that approach the boundary between categories. The results supported this hypothesis (Figure 5). Although the actual physical acoustic differences between stimuli were equal, the perceived distance was clearly *reduced* near the prototype, and *expanded* in the region of the boundary between categories. The results suggested that language experience *warped* physical space to produce a perceptual space in which perceived distances were altered. Good stimuli act like perceptual magnets by drawing tokens toward them in perceptual space. Near the boundary between two categories, the perceptual space appears to be stretched. This results in the creation of a "perceptual map" that specifies the perceptual distances and thus the relationships among stimuli. The map helps define speech categories by creating a cluster in the centre of the category and gaps at the boundaries between categories.

Development of Prototype Magnet Effects

The perceptual magnet effect was shown to be powerful for adults. Would infants demonstrate the perceptual magnet effect?

Kuhl (1991a) demonstrated the magnet effect early in life, by 6 months, prior to the time that infants uttered or understood their first words. The next question was the degree to which the magnet effect was the product of linguistic experience. Would all 6-month-old infants, regardless of language experience, exhibit the effect for the same hot spots in vowel space? Or would the effect differ in infants being reared in different language environments? If the hot spots were the same for young infants irrespective of language experience, one could argue that it constituted part of infants' innate biological endowment for language. On the other hand, the hot spots could differ in infants

Acoustic Spacing of Tokens

Perceptual Spacing of Tokens (One Dimensional MDS Solutions)

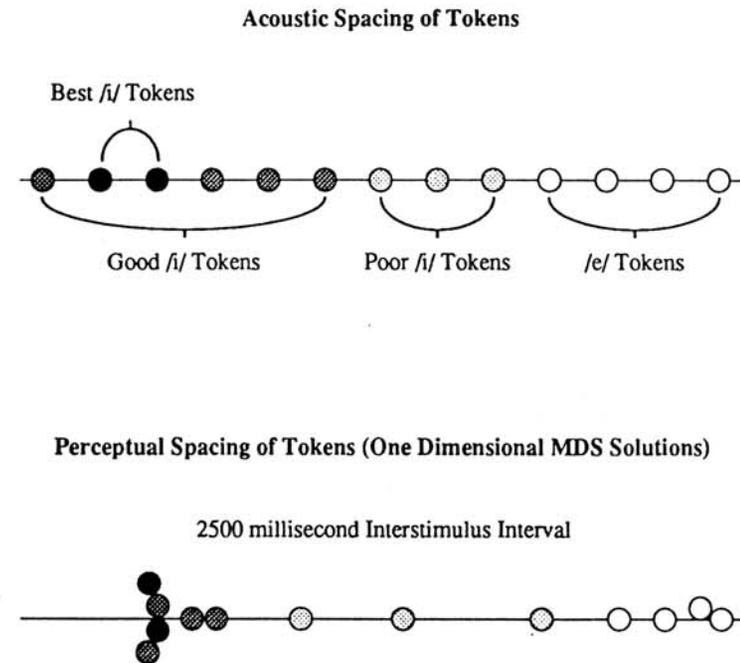**2500 millisecond Interstimulus Interval**

Figure 5: Acoustic and perceptual spacings of vowels. Vowel tokens are equidistant from one another in acoustic space; however, in perceptual space, distance is distorted. Perceptual space is shrunk in the region of the best instances of /i/ and stretched in the region of the poorest instances. From P. Iverson and P.K. Kuhl (1994), Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America* 97: 553–62.

being reared in different language environments. By 6 months of age, infants have heard a considerable amount of native-language input, and this might alter perception.

The two alternatives were tested by conducting a cross-language experiment involving English and Swedish and using vowel prototypes from both languages (Kuhl et al. 1992). Swedish was an ideal language to test the hypothesis. Infants in Sweden hear naturally occuring speech that includes three different high-front vowels, none of which is identical to American English /i/. The Swedish vowel we chose to test was the front rounded /y/, a vowel that is not produced by American adults and is thus never heard by American babies. The Swedish /y/ prototype vowel and its 32 variants were synthesized using the same techniques used to create the American English /i/ and its variants. The entire laboratory and the research team travelled to Stockholm, Sweden.

This ensured that all aspects of the tests in the two countries were identical except the language experience of the infants who were tested.

The results clearly showed that the perceptual magnet effect in 6-month-old infants was affected by exposure to a particular language. American infants demonstrated the perceptual magnet effect for the American English /i/; they treated the Swedish /y/ as a non-prototype. Swedish infants showed the opposite pattern. They demonstrated the perceptual magnet effect for the Swedish /y/ and treated the American English /i/ as a nonprototype. The data indicate a strong interaction between language environment of the infant and the sound tested. No other effects were significant. The results demonstrated an effect of language experience that was measurable by 6 months of age, clearly demonstrating that infants' exposure to ambient language alters their perception of language. This is the earliest age at which experience has been found to affect phonetic perception.

Language Experience and the Formation of Memory Representations for Speech

Kuhl (1991a, 1992a, 1993a, 1993c, 1994) argued that infants listening to language form representations of speech, creating some type of memory for the sounds of their native language. The kind of learning and memory described here is not conscious learning of specific facts or events. It could not be described as explicit, "declarative" memory (Sherry and Schacter 1987; Squire 1987; Tulving 1983, 1985). The kind of learning and memory demonstrated by infants who learn from listening to ambient language is unconscious, automatic and not due to extrinsic reinforcement; it is probably best thought of as non-declarative memory of some (as yet undefined) type. Although information about the nature of declarative memory and the brain mechanisms that control it is rapidly increasing (Squire 1987), much less is known about non-declarative memory. Such memory is likely to be species-typical and relatively permanent. It might be implicated in the type of permanent changes involved in producing an indelible "accent" or in hearing foreign-language contrasts.

If memory representations are being created as infants listen to speech, two issues will need to be addressed in future studies. Both have to do with the amount of detail preserved in speech representations: (a) do speech representations consist of individual exemplars or abstract summaries? and (b) how are the effects of speech context reflected in the representations?

Considering the first issue, early theorists assumed that because representative instances (prototypes) of categories were associated with special effects, this meant that people mentally calculated and stored some abstract version that characterized the category as a whole (Posner and Keele 1968). It was thought that perhaps an average of all the experienced exemplars was derived. An alternative, "exemplar-based" model of categorization has recently gained support (Estes 1993; Hintzman 1986; Medin and Barsalou 1987; Nosofsky 1987). According to this model, classification and the effects of good stimuli on perception can be accounted for by the storage and retrieval of individual exemplars. Exemplar theories maintain that newly encountered items act as retrieval cues to access stored individual exemplars from a category. Since the most representative (prototypic) stimuli are similar to a large number of individual exemplars, they are more likely to be accessed quickly. Thus the exemplar model offers an alternative explanation for the results of studies showing superior or more efficient recognition of prototypic items from a category.

As Estes (1993) and others have pointed out, both models account for prototype effects. In the case of speech it is not yet clear what form the underlying representation of phonetic categories might take. The magnet effect is compatible with either type of representation. Speech category information might be stored in terms of an abstract summary or as individual instances (see Kuhl 1993b, 1993c for further discussion). We underscore an additional point: there is nothing that precludes people from having access to both kinds of memory systems – one that stores information about individual exemplars and also a system that stores general category information that is derived from individual exemplars (see, e.g., Knowlton and Squire 1993).

Another issue with regard to representation is the effect of context. There are data to suggest that the location of best instances of the category shifts with changes in variables such as the rate of speech (Miller and Volaitis 1989; Volaitis and Miller 1992; Miller 1994). Similarly, we would expect that the location of the best instance of /i/ would shift with the gender of the speaker. What we do not yet know is whether a good instance produced by a male talker has an effect on perception of instances spoken by a female talker. Is the magnet's attractor effect restricted to variants that share basic parameters (such as the gender of the speaker) with the tested stimulus, or does it extend to tokens in which these basic parameters have been changed? If the representation is talker-neutral, as suggested by infants' perception of constancy for the speech of different talkers, one would expect the magnet's attractor effect to generalize.

Memory Representations for Speech Are Polymodal

The discussion thus far has concerned auditory perception. However, we do not think that speech representations are unimodal, nor do we

think that they are confined to perception. Our hypothesis is that speech representations – prototypes and the magnet effects they cause – are *polymodally mapped*, that is, they are defined in such a way that multiple sensory and motor systems have access to them. Data in support of this view come from both perception and production studies.

Perception

It was classically thought that the speech we perceived was based solely on the auditory information that reached our ears. This belief has been deeply shaken by data showing that speech perception is an intermodal phenomenon in which vision plays a role in determining what a subject reports hearing. Visual information contributes to speech perception even in the absence of a hearing impairment and even when the auditory signal is perfectly intelligible. In fact, it appears that when it is available, visual information cannot be ignored by the listener; it is automatically taken into account.

One of the most compelling examples of the polymodal nature of speech are auditory-visual "illusions" that result when discrepant information is sent to two separate modalities. Subjects report perceiving a syllable that is halfway between the one sent to the auditory and the visual system (McGurk and MacDonald 1976; Green and Kuhl 1989, 1991; Green, Kuhl, Meltzoff and Stevens 1991; Kuhl et al. 1994; Massaro 1987a, b; Summerfield 1979). One such illusion can be demonstrated when auditory information for /b/ is combined with visual information for /g/. Perceivers report the phenomenal impression of /d/ despite the fact that this information was not delivered to either sense modality.

The effect is robust and mandatory, even when a situation is created in which the discrepant information is derived from two clearly different talkers. We created a situation in which there was an obvious discrepancy between the gender of the talker presenting the information in the two modalities (Green et al. 1991). A male face was combined visually with the voice of a female talker, and vice versa. We took pains to choose our speakers such that the gender incompatibility was highly salient. A very male-looking football player's face was paired with a high and feminine-sounding female voice, and vice versa. There was no mistaking the gender mismatch.

The results showed that even though the gender discrepancy was readily apparent to viewers, they nonetheless integrated the auditory and visual information, reporting that they perceived the illusory consonant /d/. The effect was as pervasive in the gender-discrepancy situation as it was when the gender of the talker remained constant. The results show that even in situations in which the two inputs could not have derived from a common biological source, the integration of the

information from the two modalities is not disrupted. Observers knew that the two inputs did not go together; yet they were compelled to integrate them. This demonstrates how thoroughly speech is polymodally specified.

Even very young infants appear to represent speech polymodally. Infants reveal their knowledge in two situations: when watching and listening to another person speak, and when attempting to imitate a sound they hear another produce. We demonstrated that 18- to 20-week-old infants recognize auditory-visual connections, akin to what we as adults do when we lipread (Kuhl and Meltzoff 1982, 1984; Kuhl, Williams and Meltzoff 1991). In the experiment, infants viewed two filmed faces, side by side, of a woman pronouncing two vowels silently, the vowel /a/ and the vowel /i/ (Figure 6). The faces pronouncing the
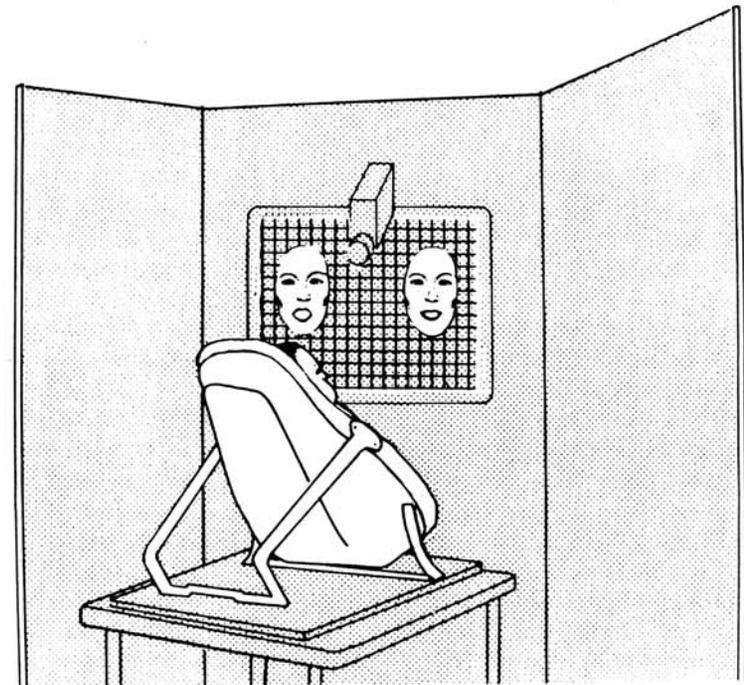

Figure 6: Technique used to test infant cross-modal (auditory-visual) speech perception. Infants watched two faces, side-by-side, producing two different vowels, /a/ and /i/. At the same time they listened to a single vowel (either /a/ or /i/) presented from a loudspeaker located midway between the two faces. The results demonstrated that 18- to 20-week-old infants looked longer at the face that matched the vowel they heard. From P.K. Kuhl and A.N. Meltzoff (1982), The bimodal perception of speech in infancy. *Science* 218: 1138–41.

two vowels opened and closed in perfect synchrony; one mouthed the vowel /a/ and the other the vowel /i/. While viewing the two faces, infants heard one of the two vowels (either /a/ or /i/), played from a loudspeaker located midway between the two faces. The sound was played in synchrony with the two facial movements. The results of the test showed that infants who heard the vowel /a/ looked longer at the face pronouncing /a/, while the infants who heard the vowel /i/ looked longer at the face pronouncing /i/. The only way infants could do this is by recognizing the correspondence between the auditory and visual speech information – there were no temporal or spatial clues telling the infant which face uttered the sound. The experiment shows that by 4 months, infants appear to know that an /a/ sound "goes with" a face mouthing /a/, while an /i/ vowel "goes with" a face mouthing /i/. Infants' cross-modal speech perception abilities indicate that they are beginning to recognize the relationship between sound and articulatory movement, at least when they observe another person speak. (See also MacKain, Studdert-Kennedy, Spieker and Stern 1983; Walton and Bower 1993.)

Production

Can infants relate sound to movement in their own speech? Studies on vocal imitation offer evidence that infants hearing an auditory signal know what to do with their own articulators to produce the sounds themselves. In a recent experiment, infants' vocalizations in response to speech were recorded at three ages – at 12, 16 and 20 weeks (Kuhl and Meltzoff, in press). Infants listened to a woman producing one of three vowels, /a/, /i/, or /u/ . Infants' vocalizations were analyzed perceptually by having them phonetically transcribed, and analyzed instrumentally using computerized spectrographic techniques.

Two findings are noteworthy. First, there was developmental change in infants' vowel productions. Figure 7 displays the vowels of 12-, 16- and 20-week-old infants in an acoustic space. In each graph, infants' vowel utterances are classified according to a transcription provided by a phonetically trained listener. The closed circles enclose 90% or more of the utterances in each category. As shown, utterances in each of the three categories formed clusters in acoustic space. More importantly, the areas of vowel space occupied by infants' /a/, /i/, and /u/ vowels become progressively more *separated* between 12 and 20 weeks of age. Infant vowel categories were more tightly clustered at 20 weeks than at 12 weeks. What causes the increased separation of vowel categories over this relatively short (8-week) period? We suggest that infants listening to their ambient language have begun to form *memory representations* of vowels; these representations serve as "targets" that infants
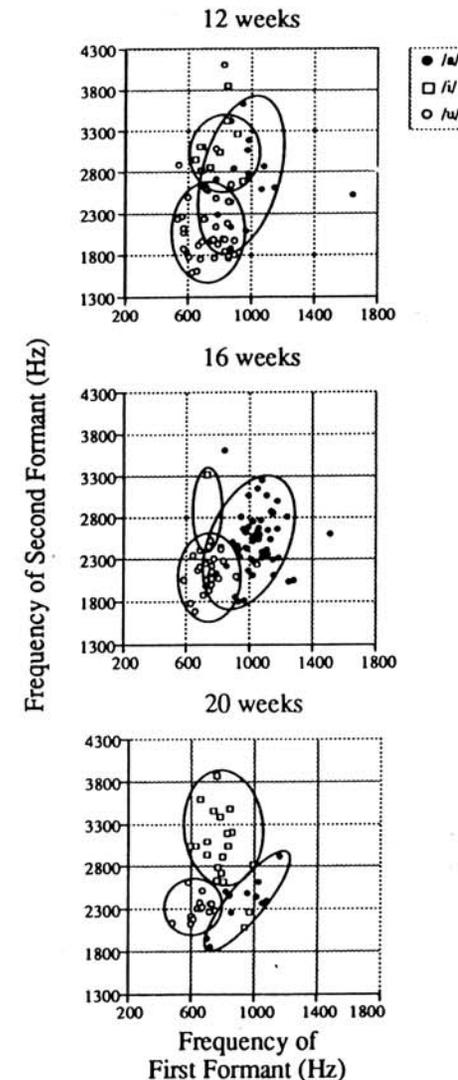

Figure 7: The location of /a/, /i/, and /u/ vowels produced by 12-, 16- and 20-week-old infants. The curves were drawn by visual inspection to enclose 90% or more of the infants' utterances. Across age, infants' vowel productions show tighter clustering in vowel space. From P.K. Kuhl and A.N. Meltzoff (in press), Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America.*

try to match. Our view is that the memory representations resulting from infants' exporsure to ambient language influences not only their perception of speech but their subsequent productions as well.

A second result of the study demonstrated that infants' vowel productions can be influenced by short-term exposure to sound. Infants' vowel productions were influenced by what they heard. Infants produced more /a/-like utterances when exposed to /a/ than when exposed to /i/ or /u/; similarly, they produced more /i/-like utterances when exposed to /i/ than when exposed to /a/ or /u/; finally, they produced more /u/-like utterances when exposed to /u/ than when exposed to /a/ or /i/. In short, infants vocally imitated the gross spectral quality of the stimulus they heard.

The surprising thing was that the total amount of exposure that infants received was only 15 minutes (5-min. exposure to a specific vowel for each of three days). If 15 minutes of laboratory exposure to a vowel is sufficient to influence infants' vocalizations, then listening to the ambient language for 12 weeks certainly provides sufficient exposure to induce change. There is some evidence from the results of babbling studies conducted on infants from different cultures that 1-year-olds in different cultures have begun to be influenced by native-language input (de Boysson-Bardies, Sagart and Durand 1984; de Boysson-Bardies, Halle, Sagart and Durand 1989). The new experimental data demonstrate that even short-term laboratory exposure is sufficient to alter infants' vocal productions.

How do 12-week-old infants know how to move their articulators in a way that achieves a specific auditory target? Some primitive tendency for human speech to drive infants' vocal productions may exist that is analogous to the innate tendency for visually perceived body movements to drive corresponding motor acts, as manifest in newborn gestural imitation (Meltzoff and Moore 1977, 1983, 1992, 1994). Meltzoff and Moore have shown that in the absence of sound, infants imitate movements that involve the speech articulators, such as mouth opening and tongue protrusion (Meltzoff and Moore 1977). The youngest infant tested in this work was 42 minutes old. This is truly an innate ability, one documented in the first *hour* of postnatal life (Meltzoff and Moore 1983). We do not know if an innate mapping from auditory to articulatory events exists, but it is not out of the question, given Meltzoff and Moore's findings of visual-motor mappings of mouth movements used in speech.

Even if primitive connections exist initially, they must be rapidly expanded to create the repertoire that infants possess just a short time later. This rapid expansion is gained, we believe, through experience as infants engage in cooing and sound play. Infant cooing, which begins

at about 4 weeks of age, allows extensive exploration of the nascent auditory-articulatory map during which (self-produced) auditory events are related to the motor movements that caused them. Presumably, infants' accuracy in producing vowels improves as infants relate the acoustic consequences of their own articulatory acts to the acoustic targets they heard. This account implies that infants not only have to be able to hear the sounds produced by others, but that they need to hear the results of their own attempts to speak in order to make progress. Both hearing the sound patterns of ambient language (auditory exteroception) and being able to hear one's own attempts at speech (auditory proprioception) are critical determinants of vocal development.

Social Context: Motherese and the Effects of Language Input

The cross-language perception results, coupled with the developmental change in vocal production, suggest that as infants listen to early speech their perception-motor system is being altered. Infants are bathed in language from the time they are born, and this early language experience affects them. What do we know about the nature of early linguistic input?

We know that the prosodic characteristics of "motherese" are unique: it has a higher pitch, a slower tempo and exaggerated intonation contours (Fernald and Simon 1984). We also know that this speaking style is near universal in the speech of caretakers around the world when addressing infants (Grieser and Kuhl 1988). Motherese is socially pleasing and attention-getting, and parents from almost all cultures use it when speaking to their infants. Research has also shown that infants prefer to listen to motherese over speech that is directed towards another adult (Fernald 1985; Fernald and Kuhl 1987). Motherese attracts infants' attention. Are the phonetic units contained in motherese somehow special?

Motherese is "vowel-drenched," and the vowels contained in motherese tend to be prolonged due to its slower tempo. A recent study in Kuhl's laboratory examined the phonetic content of motherese. Women were recorded speaking naturally to their 2-month-old infants and to an adult. They were told to use three words containing the vowel /i/ in both conversations: "bead," "keys," and "sheep." The three words were edited out of these dialogues and rated by adults using the 7-point goodness rating scale. The study revealed that the vowels contained in motherese were perceived as better instances than the same vowels spoken by the same women when addressing an adult. The vowels of motherese may thus be ideal signals for learning.

Are the higher pitch and expanded intonation contours typical of motherese *necessary* for learning? Our guess is that they are not. The context in which language is presented to the child – both the auditory

characteristics of motherese and its visual aspects (greatly exaggerated facial expressions) – grab infant attention and fix it on the talking caretaker. This might maximize learning but exaggerated pitch and intonation are probably not necessary for language learning. It would be useful to examine whether interaction with another human being is necessary for this kind of learning to occur. Infants have been shown innately to react to other humans (Meltzoff and Moore 1983, 1993, 1995), and learning often occurs in interpersonal encounters. A person may be a biological signal that triggers the kind of learning we have described. Nonetheless, we would be interested in knowing whether infants can learn language information in the absence of a person. Would, for example, an infant's perception of speech be altered by playing Berlitz language tapes via a device hung on the infant's crib.

Evolution: Are Language Precursors Uniquely Human?

Language in human infants, even at the level of phonetics, has been shown to have multiple determinants: innate perceptual predispositions; magnet effects that alter perceptual space; a cognitive system that forms memory representations accessible to multiple modalities and (possibly) social interaction among people. Is this entire set of abilities uniquely human? Are any aspects of this composite common to humans and other animals? Modern studies of speech perception reveal that some of the speech effects found in human infants can be replicated by a monkey (see Kuhl 1991b for summary). Others appear to be uniquely human. The relevant studies have been done with three different phenomena – CP, speech prototypes and auditory-vocal mapping – and the points of convergence and divergence between monkeys and humans is of considerable interest.

Tests of CP were conducted by Kuhl and colleagues on animals whose auditory systems are very similar to humans', such as chinchillas and monkeys (Kuhl and Padden 1982, 1983; Kuhl and Miller 1975). The results showed that animals responded as though they heard a sudden change in the speech stimuli at the exact location where human adults perceived a shift from one phonetic category to another.

These findings on animals influenced our theories of infants. It showed that the innate CP effects in human infants are not, in themselves, evidence compelling the postulation of a speech module. Evidently, CP can be accomplished in the absence of a speech module, because animals also show the same CP effects as human newborns. Kuhl theorized that CP in infants – the perception of "basic cuts" – was attributable to a *general auditory processing mechanism* rather than a special language module, and that it was very deeply embedded in our phylogenetic history (Kuhl and Miller 1975, 1978; Kuhl 1988). On this

view, the perception of basic cuts in auditory signals, which would have been available in early hominids, was exploited in the evolution of the sound system used in language (Kuhl 1987, 1991b). It may have helped in determining which oral sounds were good candidates for language to use.

In contrast to CP, humans and animals strongly diverge in tests examining the prototype's magnet effect (Kuhl 1991a). Monkeys displayed no magnet effect; they equated variants to the prototype and the non-prototype to the same degree. Whether or not animals would learn speech prototypes if they were repeatedly exposed to speech in a social setting is an interesting question. We doubt that the kind of learning that we have described for speech would take place in monkeys. It would be even less likely to take place if we placed a tape recorder playing the sounds of language inside the monkey's cage. The kind of learning that we are describing may well require a social setting and interaction among conspecifics. (Monkeys would perhaps exhibit magnet effects for the perception of their own calls.) We thus offer the hypothesis that the formation of perceptual representations based on experience is innately guided and species-specific. Human infants may have to interact with other persons who are perceived as "like me" (Meltzoff and Gopnik 1993; Meltzoff and Moore 1995) before this kind of learning is triggered.

Is the cross-modal representation of speech information species-specific? Non-human primates may lack the cross-modal connections between the auditory-vocal channel necessary for both auditory-visual speech perception and for vocal imitation. *Homo sapiens* is the only mammal that displays vocal learning, the tendency to acquire the species-typical vocal repertoire by hearing the vocalizations of adults and mimicking them. Humans share this ability with a few avian species, the songbirds (Marler 1976; Konishi 1989), who learn their species-specific songs if they are exposed to them during a sensitive period early in life (Nottebohm 1975). In the case of birds there are data showing that learning is enhanced in the presence of a visual instance of the species. In fact, the presence of a conspecific bird allows a young bird to learn some of the notes of an alien species. This suggests how intricately woven the mechanisms of learning may be with social (hormonal?) aspects.

A Theory of Early Speech Development

The diverse research described here has been integrated into a model of the development of speech perception called the *Native Language Magnet* (NLM) theory (Kuhl 1992a, 1993a, 1993b, 1994). The theory encompasses the initial state as well as the changes brought about by

experience with language. It explains how infants' developing speech representations alter both speech perception and speech production.

NLM theory holds that what is "given by nature" is the ability to partition the sound stream into gross categories separated by natural boundaries. As illustrated in Figure 8, these boundaries are what is tested in studies of CP. Tests of CP have shown that infants are sensitive to the acoustic cues that underlie phonetic distinctions (both consonants and vowels) in language. Infants' abilities to partition the acoustic stream serves to *initially structure* phonetic perception.

The boundary effects associated with CP are also displayed by non-human animals. Thus, perceptual boundaries are not due to an innate module that evolved for language. Infants' abilities to hear the relevant differences between phonetic units is innate, but is based on a long phylogenetic history and attributable to general auditory processing mechanisms. These general auditory processing abilities are neither language-specific nor species-specific.

By 6 months, infants have something more than the innate "basic cuts" for phonetic perception. By 6 months, infants show *language-specific* magnet effects, due to memory representations of the sound patterns of the ambient language. The development of magnet effects is illustrated in Figure 9. Here magnet effects are schematically presented for infants being raised in Sweden, America and Japan. Both the number and location of vowel sounds differ across the three languages. The graphs are not meant to precisely mark the locations of vowel magnets

Figure 9: By 6 months of age, infants reared in different linguistic environments show an effect of language experience. They exhibit language-specific magnet effects that result from listening to the ambient language. From P.K. Kuhl (1994), Learning and representation in speech and language, *Current Opinion in Neurobiology* 4: 812–22.

but to convey in conceptual terms that linguistic experience in the three different cultures has resulted in magnet effects (memory representations with magnet-like properties) that differ in number and location for infants growing up listening to the three different languages.

Magnet effects are the result of infants' stored representations of language input. These representations are derived from infants' analysis of the distributional properties of speech produced by native speakers of the language, mostly but not exclusively in interaction with them. Infants' analysis of ambient language results in learning of the properties of the native language; as a result, infants commit a particular language's regularities to memory. Infants' initial perceptual boundaries assist this process: boundaries set limits on the area that infants' representations must organize. Because of this, infants do not form representations that encompass the entire vowel space; instead, infants' representations organize input that falls within a bounded area of the vowel space. The innate partitioning (Figure 8) thus constrains and helps organize language input to the child.
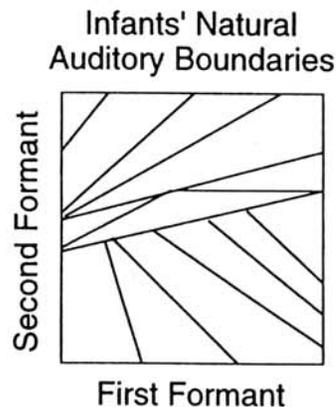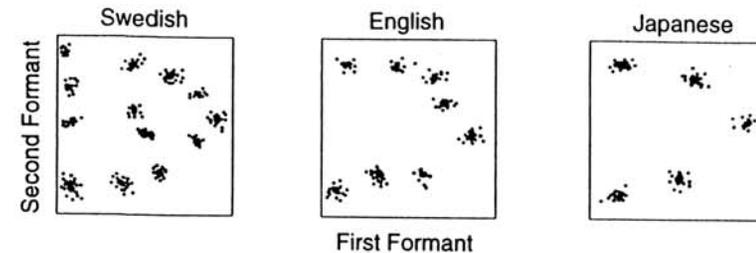

Figure 8: At birth, infants perceptually partition the acoustic space underlying phonetic distinctions in a language-universal way. They are capable of discriminating all phonetically relevant differences in the world's languages. From P.K. Kuhl (1994), Learning and representation in speech and language, *Current Opinion in Neurobiology* 4: 812–22.

Effects of Speech Representations on Speech Perception

What effects do infants' stored representations (shown in Figure 9) have on speech perception? According to NLM, of speech stored in memory representations are responsible for the magnet effect observed in experiments. However, infants' stored representations go beyond this and affect the perception of foreign-language sounds as well. The warping of acoustic space causes certain perceptual distinctions to be minimized (those near the magnet attractors) while others are maximized (those near the boundaries between two magnets). The result is that some of the boundaries that innately divided the space "disappear" as the

perceptual space is reconfigured to incorporate a language's particular magnet placement. This is schematically illustrated in the diagrams of Figure 10. Magnet effects functionally erase certain boundaries – those relevant to foreign but not native languages.[1]

NLM thus offers an explanation for two related findings: (a) infants' loss of discrimination of foreign-language contrasts they once had (Werker 1991), and (b) adult reactions to foreign sounds. According to NLM, the developing magnet pulls sounds that were once discriminable towards a single magnet, making them no longer discriminable. The prediction is that magnet effects occur first, before the failure to discriminate; they developmentally precede and underlie the changes in infants' perception of foreign-language contrasts. They thus offer a *mechanism* that explains the change in phonetic perception that Werker observed. The magnet effect also helps account for the results of studies on the perception of sounds from a foreign language by adults. These studies suggest that phonetic units from a foreign language that are similar to a category in the adult's own native language are particularly difficult to perceive as different from the native-language sound; sounds not similar to a native-language category are relatively easy to discriminate (Best 1993; Best, McRoberts and Sithole 1988). The theory accounts for this because the prototype of, for example, the Japanese category is similar to both /r/ and /l/; its magnet effect makes the two sounds difficult for native-speaking Japanese people to discriminate. The prediction from this theory is that the difficulty posed by a given foreign-language unit will depend on its proximity to a native-language magnet. The nearer it is to a magnet, the more it will be pulled toward the native-language category, making it indistinguishable from the native-language sound. One way to visualize the effects of native-language magnet effects on foreign-language sounds is to imagine that native-language representations serve as filters for in-
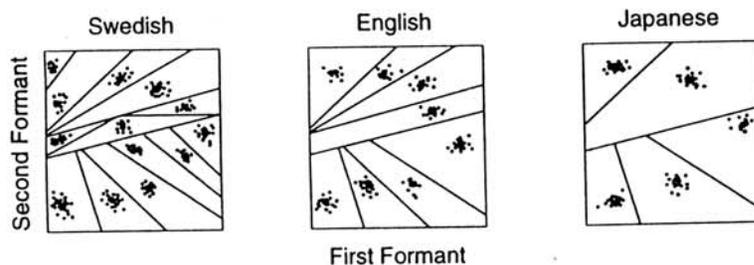

Figure 10: After language-specific magnet effects appear, certain phonetic boundaries are functionally "erased." From P.K. Kuhl (1994), Learning and representation in speech and language, *Current Opinion in Neurobiology* 4: 812–22.

coming sounds; the sounds of a second language have to be "pulled through" the filters formed by the first.

Effects of Speech Representations on Production

Human infants learn speech by listening to ambient language and attempting to produce sound patterns that match what they hear. The specific inventory of phonetic units, words and prosodic features employed by a particular language are learned largely through imitation (broadly construed). By 2 years of age, infants have begun to "sound like" a native speaker of their language.

The theory developed here attributes infants' learning of the speech patterns to the memory representations formed in early infancy by listening to ambient language. The speech patterns stored in memory serve as "targets" that infants try to match. The data gathered in our study of infant imitation (Kuhl and Meltzoff, in press) provided evidence of long-term change in infants' vocal repertoires, changes that occurred over the 8-week period (from 12 to 20 weeks) during which infants' vocalizations were measured (Figure 7), as well as evidence of short-term change in infants' vocalizations, changes that occurred with 15 minutes of laboratory exposure. These findings strongly suggest that infants' acquisition and production of speech is highly influenced by the auditory information that surrounds them. Speech representations learned in early infancy drive infants' early production of speech and account for the fact that adult speakers produce speech with an "accent." Hearing a specific language early in life puts an indelible mark on one's speech.

Two streams of research – that linguistic exposure alters infants' perception of speech (Kuhl et al. 1992) and linguistic exposure alters infants' production of speech (Kuhl and Meltzoff, in press) – can thus be unified by the suggestion that memory representations of speech underlie both findings. The emerging view suggests that the tighter clustering observed in infant vowel production and the tighter clustering among vowels in infant perception are both attributable to a common underlying mechanism – the formation of memory representations that derive initially from perception of the ambient input and then act as targets for motor output (Figure 11). The speech representational system is thus deeply and thoroughly polymodal. Early experience affects both sensory perception and motor learning.

Summary and Conclusions

We have described a three-step theory of speech development (Native Language Magnet theory or NLM), which embraces nativism, evolution and learning in the development of speech.
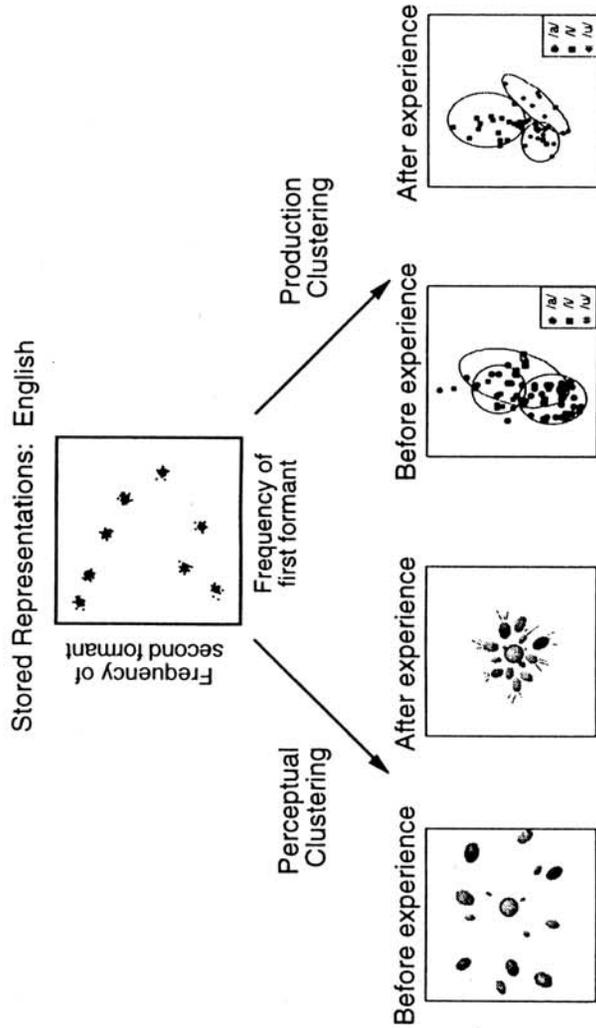
Figure 11: Infants' stored representations affect both speech perception, producing the perceptual clustering evidenced by the magnet effect, as well as speech production, producing the increased clustering seen in infants' vocalizations. From P.K. Kuhl (1994), Learning and representation in speech and language. *Current Opinion in Neurobiology* 4: 812–22.

Native Language Magnet Theory

Phase 1: Infants are born with innate boundaries that partition the incoming speech stream into phonetically relevant categories. Innate perceptual abilities undergird language learning. Auditory boundaries deeply embedded in our phylogenetic history strongly influenced the original selection of sounds used in the world's languages. Infants' abilities are thus derived from a mechanism that is innate and attributable to a general auditory processing mechanism (not a language module per se).

Phase 2: Exposure to ambient language results in speech representations, stored memories of the sound patterns that reflect the distributional properties of the infant's native language. Infants' stored representations of speech have unique psychological properties and indicate a special kind of learning. Regarding the psychological properties, two facts stand out. First, stored representations produce perceptual magnet effects: members of the category are clustered in perceptual space. Thus we say that perceptual magnets *warp* the underlying space. Second, stored representations are polymodal, coding not only the auditory properties of language but its visual and motor aspects as well. Stored representations thus contain information from a variety of modalities. Regarding learning, the kind of learning described here is not the typical stimulus-response learning with external reinforcement that psychologists often describe. It is unconscious, automatic, long-lasting and extremely difficult to undo. This kind of learning is likely to be species-specific and highly constrained.

Phase 3: Perceptual space is reconfigured such that certain innate perceptual boundaries have been functionally erased. The theory is recursive in the sense that the output of phase 2 reaches back and reorganizes the innately given framework. A space characterized by simple boundaries has been replaced by a dynamically warped space dominated by magnets.

What NLM Theory Accounts For

The Native Language Magnet theory accounts for a wide variety of facts in the perception and production of speech by infants and adults. Some of these are the following.

(a) The loss of discrimination that was innately present. Infants' language-general speech perception abilities change to ones that are language-specific. By the end of the first year, infants fail to discriminate contrasts that they once demonstrated the ability to discriminate. According to the theory, this change is due to the development of speech representations and their corresponding magnet effects.

(b) Speech production converges on target sounds of the native language. Infants' language-general speech production capabilities change to ones that are language-specific. According to the theory, the production system is guided by infants' speech representations. Infants babble, and in doing so they provide themselves with the opportunity to compare the auditory results of their own productions to the representations stored in memory. This sound play also enriches the sensory-motor map specifying the relationship between mouth movements and sound. The map helps infants correct their motor movements to converge on the intended target, which is specified in the memory representation.

(c) Second-language learning is difficult for adults. Memory representations for speech are formed early during a period in which we are particularly sensitive to language experience. These representations and the magnet effects they cause can be thought of as forming a set of filters through which all subsequent language input passes. The set of filters established when we learn our primary language subsequently interferes with the ability to acquire a second language. This is because the filters required to process the second language do not match those that characterize the first.

(d) "Accent." Speakers of different languages, and even speakers of the same language who speak different dialects, exhibit an "accent," a speaking pattern that identifies them as coming from a particular language group. Accents involve motor patterns that are unconsciously learned. Accents indelibly mark our speech even after we acquire a new language, providing evidence of interference between the production pattern of one's primary language and the production pattern of a new language.

(e) Early learning is long-lasting and difficult to alter. The effects on speech perception and speech production are long-lasting and difficult to change. The acquisition of a second language past puberty is more difficult than at an earlier age. This is also true for the elimination of an accent. Our native language indelibly marks us through the formation of representations that "distort" future inputs but are not easily changed by them. Once acquired, the altered magnet-induced perceptual map and the magnet-driven production patterns that characterize our native language resist change.

(f) Polymodal effects: Visual speech perception is mandatory. If infants' stored representations of speech have encoded both the auditory and visual aspects of speech, then one would expect that when visual information accompanies speech, observers would be compelled to take

that information into account. This result has been confirmed in numerous experiments on auditory-visual speech perception in adults. The fact that infants of only 18 weeks demonstrate an ability to relate auditory speech to its visual concomitants further supports the view that the representations are polymodal in nature right from the earliest phases of life.

Final Thoughts

Our research shows that infants' innate abilities initially structure language growth. In that sense we are nativists. However, these innate abilities do not solely determine language. The three-step model we described shows how linguistic experience yields representations that recursively alter the innately provided system in a profound and long-lasting way.

The type of learning we described does not correspond to Skinner's conditioning and extrinsic reinforcement. Nor is the restructuring well captured by the notion that "learning" is simple triggering nor by the idea that "learning" is the deletion or subtraction of information that was innately present. It is true that perceptual boundaries are innately determined and then deleted (in a sense), but the deletion of boundaries would not, by itself, produce the phonology of a particular language. The critical information for a mature phonology is not where the boundaries are located. The critical information for phonology is where in phonetic space a particular language locates its category centres and the distribution of members around those centres. Particular languages utilize only some of the bounded spaces and place category prototypes in different locations within bounded spaces. The functional erasure of boundaries is thus a secondary process caused by the formation of phonetic category centers, hot spots in space that vary as a function of ambient language input. It is the formation of category centres and the perceptual similarity space that surrounds them that needs to be understood in seeking the developmental roots of a particular language's phonology.

The learning we have described in speech development – unconscious, automatic and long lasting – is species-typical and probably species-specific. The learning itself may have two biological constraints. It may require social interaction with other humans – the recognition of other humans itself being innately specified (Meltzoff and Moore 1995). For this reason, infants' perception may not be altered by playing Berlitz language tapes. A further constraint is provided by a neural system prepared to receive input at a particular time in ontogenesis, described by Greenough as an "experience-expectant" neural process (Greenough and Alcantara 1993; Greenough and Black 1992).

Greenough argues that neural systems are prepared for experience by an overproduction of synaptic connections that are subsequently pruned to achieve a more efficient neural organization. Synaptic overproduction occurs for cases in which a specific kind of experience is highly reliable in the environment of the organism. Language input is a reliable feature of infants' early postnatal growth. Thus, evolution, nativism and experience all meet in the human infant. The indelible mark of early speech or language learning may be traceable to a biological preparation for receiving language input from other humans during the first years of life.

Acknowledgment

Note

1 Work on adults suggests that the boundaries do not literally disappear; with training it is possible to increase performance on the discrimination of foreign-language contrasts in adults (e.g., MacKain, Best and Strange 1981; Logan, Lively and Pisoni 1991). Thus the alterations that occur do not involve changes at a sensory level, but ones at a higher level involving memory and/or attention.

References

Abramson, A.S., and L. Lisker (1970). Discriminability along the voicing continuum: Cross-language tests. In *Proceedings of the Sixth International Congress of Phonetic Sciences Prague 1967*, 569–73. Prague: Academia

Best, C.T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage and J. Morton (eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, 289–304. Dordrecht: Kluwer

Best, C.T., G.W. McRoberts and N.M. Sithole (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14: 345–60

Davis, K., and P.K. Kuhl (1992). Best exemplars of English velar stops: A first report. In J.J. Ohala, T.M. Nearey, B.L. Derwing, M.M. Hodge and G.E. Wiebe (eds.), *Proceedings of the International Conference on Spoken Language Processing*, (495–98). Edmonton, AB: University of Alberta

——— (1993). Acoustic correlates of phonetic prototypes: Velar stops. *Journal of the Acoustical Society of America* 93: 2392

de Boysson-Bardies, B., P. Halle, L. Sagart and C. Durand (1989). A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language* 16: 1–17

de Boysson-Bardies, B., L. Sagart and C. Durand (1984). Discernible differences in the babbling of infants according to target language. *Journal of Child Language* 11: 1–15

Eimas, P.D., E.R. Siqueland, P. Jusczyk and J. Vigorito (1971). Speech perception in infants. *Science* 171: 303–6

Estes, W.K. (1993). Concepts, categories, and psychological science. *Psychological Science* 4: 143–53

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development* 8: 181–95

Fernald, A., and P. Kuhl (1987). Acoustic determinants of infant preference for Motherese speech. *Infant Behavior and Development* 10: 279–93

Fernald, A., and T. Simon (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology* 20: 104–13

Fodor, J.A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology.* Cambridge, MA: MIT Press

Green, K.P., and P.K. Kuhl (1989). The role of visual information in the processing of place and manner features in speech perception. *Perception and Psychophysics* 45: 34–42

——— (1991). Integral processing of visual place and auditory voicing information during phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance* 17: 278–88

Green, K.P., P.K. Kuhl, A.N. Meltzoff and E.B. Stevens (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. *Perception and Psychophysics* 50: 524–36

Greenough, W.T., and A.A. Alcantara (1993). The roles of experience in different developmental information stage processes. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, and J. Morton (eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, 3–16. Dordrecht: Kluwer

Greenough, W.T., and J.E. Black (1992). Induction of brain structure by experience: Substrates for cognitive development. In M. Gunnar and C. Nelson (eds.), *The Minnesota Symposia on Child Psychology, Vol. 24: Developmental Behavioral Neuroscience*, 155–200. Hillsdale, NJ: Erlbaum

Grieser, D.L., and P.K. Kuhl (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology* 24: 14–20

——— (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology* 25: 577–88

Hintzman, D.L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review* 93: 411–28

Iverson, P., and P.K. Kuhl (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America* 97: 553–62

—— (in press). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *Journal of the Acoustical Society of America.*

Knowlton, B.J., and L.R. Squire (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science* 262: 1747–49

Konishi, M. (1989). Birdsong for neurobiologists. *Neuron* 3: 541–49

Kruskal, J.B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29: 1–27

Kuhl, P.K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America* 66: 1668–79

—— (1987). The special-mechanisms debate in speech research: Categorization tests on animals and infants. In S. Harnad (ed.), *Categorical Perception: The Groundwork of Cognition*, 355–86. New York: Cambridge University Press

—— (1988). Auditory perception and the evolution of speech. *Human Evolution* 3: 19–43

—— (1991a). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics* 50: 93–107

—— (1991b). Perception, cognition, and the ontogenetic and phylogenetic emergence of human speech. In S.E. Brauth, W.S. Hall, and R.J. Dooling (eds.), *Plasticity of Development*, 73–106. Cambridge, MA: MIT Press

—— (1992a). Infants' perception and representation of speech: Development of a new theory. In J.J. Ohala, T.M. Nearey, B.L. Derwing, M.M. Hodge and G.E. Wiebe (eds.), *Proceedings of the International Conference on Spoken Language Processing*, 449–56. Edmonton, AB: University of Alberta

—— (1992b). Psychoacoustics and speech perception: Internal standards, perceptual anchors, and prototypes. In L.A. Werner and E.W. Rubel (eds.), *Developmental Psychoacoustics*, 293–332. Washington, DC: American Psychological Association

—— (1993a). Innate predispositions and the effects of experience in speech perception: The Native Language Magnet theory. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, and J. Morton (eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, 259–74. Dordrecht: Kluwer

—— (1993b). Infant speech perception: A window on psycholinguistic development. *International Journal of Psycholinguistics* 9: 33–56

—— (1993c). Early linguistic experience and phonetic perception: Implications for theories of developmental speech perception. *Journal of Phonetics* 21: 125–39

—— (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology* 4: 812–22

Kuhl, P.K., and P. Iverson (1995). Linguistic experience and the "perceptual magnet effect." In W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium, MD: York Press

Kuhl, P.K., and A.N. Meltzoff (1982). The bimodal perception of speech in infancy. *Science* 218: 1138–41

—— (1984). The intermodal representation of speech in infants. *Infant Behavior and Development* 7: 361–81

—— (in press). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America*

Kuhl, P.K., and J.D. Miller (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science* 190: 69–72

—— (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America* 63: 905–17

Kuhl, P.K., and D.M. Padden (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception and Psychophysics* 32: 542–50

—— (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America* 73: 1003–10

Kuhl, P.K., M. Tsuzaki, Y. Tohkura and A.N. Meltzoff (1994). Human processing of auditory-visual information in speech perception: Potential for multimodal human-machine interfaces. In *Proceedings of the International Conference on Spoken Language Processing*, 539–42. Tokyo: Acoustical Society of Japan

Kuhl, P.K., K.A. Williams, F. Lacerda, K.N. Stevens and B. Lindblom (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255: 606–8

Kuhl, P. K., K.A. Williams and A.N. Meltzoff (1991). Cross-modal speech perception in adults and infants using nonspeech auditory stimuli. *Journal of Experimental Psychology: Human Perception and Performance* 17: 829–40

Lasky, R.E., A. Syrdal-Lasky and R.E. Klein (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology* 20: 215–25

Liberman, A.M., F.S. Cooper, D.P. Shankweiler and M. Studdert-Kennedy (1967). Perception of the speech code. *Psychological Review* 74: 431–61

Logan, J.S., S.E. Lively and D.B. Pisoni (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America* 89: 874–86

MacKain, K.S., C.T. Best and W. Strange (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics* 2: 369–90

MacKain, K., M. Studdert-Kennedy, S. Spieker and D. Stern (1983). Infant intermodal speech perception is a left-hemisphere function. *Science* 219: 1347–49

Marean, G.C., L.A. Werner and P.K. Kuhl (1992). Vowel categorization by very young infants. *Developmental Psychology* 28: 396–405

Marler, P. (1976). Sensory templates in species-specific behavior. In J.C. Fentress (ed.), *Simpler Networks and Behavior*, 314–29. Sunderland, MA: Sinauer

Massaro, D.W. (1987a). Psychophysics versus specialized processes in speech perception: An alternative perspective. In M.E.H. Schouten (ed.), *The Psychophysics of Speech Perception*, 46–65. Boston: Nijhoff

———— (1987b). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry.* Hillsdale, NJ: Erlbaum

McGurk, H., and J. MacDonald (1976). Hearing lips and seeing voices. *Nature* 264: 746–48

Medin, D.L., and L.W. Barsalou (1987). Categorization processes and categorical perception. In S. Harnad (ed.), *Categorical Perception: The Groundwork of Cognition*, 455–90. New York: Cambridge University Press

Meltzoff, A.N. (1990). Towards a developmental cognitive science: The implications of cross-modal matching and imitation for the development of representation and memory in infancy. In A. Diamond (ed.), *The Development and Neural Bases of Higher Cognitive Functions: Annals of the New York Academy of Sciences* 608: 1–31

Meltzoff, A., and A. Gopnik (1993). The role of imitation in understanding persons and developing a theory of mind. In S. Baron-Cohen, H. Tager-Flusberg and D.J. Cohen (eds.), *Understanding Other Minds: Perspectives from Autism*, 335–66. New York: Oxford University Press

Meltzoff, A.N., and M.K. Moore (1977). Imitation of facial and manual gestures by human neonates. *Science* 198: 75–78

———— (1983). Newborn infants imitate adult facial gestures. *Child Development* 54: 702–9

———— (1992). Early imitation within a functional framework: The importance of person identity, movement, and development. *Infant Behavior and Development* 15: 479–505

———— (1993). Why faces are special to infants: On connecting the attraction of faces and infants' ability for imitation and cross-modal processing. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage and J. Morton (eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, 211–25. Dordrecht: Kluwer

———— (1994). Imitation, memory, and the representation of persons. *Infant Behavior and Development* 17: 83–99

———— (1995). Infants' understanding of people and things: From body imitation to fold psychology. In J. Bermúdez, A.J. Marcel and N. Eilan (eds.), *The Body and the Self*, 43–69. Cambridge, MA: MIT Press

Mervis, C.B., and E. Rosch (1981). Categorization of natural objects. *Annual Review of Psychology* 32: 89–115

Miller, J.L. (1994). On the internal structure of phonetic categories: A progress report. *Cognition* 50: 271–85

Miller, J.L., and L.E. Volaitis (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception and Psychophysics* 46: 505–12

Miyawaki, K., W. Strange, R. Verbrugge, A.M. Liberman, J.J. Jenkins and O. Fujimura (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics* 18: 331–40

Nosofsky, R. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition* 15: 87–108

Nottebohm, F. (1975). A zoologist's view of some language phenomena with particular emphasis on vocal learning. In E.H. Lenneberg and E. Lenneberg (eds.), *Foundations of Language Development*, Vol. 1, 61–103. San Francisco: Academic Press

Posner, M.I., and S.W. Keele (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology* 77: 353–63

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology* 7: 532–47

———— (1978). Principles of categorization. In E. Rosch and B. Lloyd (eds.), *Cognition and Categorization*, 27–48. Hillsdale, NJ: Erlbaum

Rosch, E.H. (1977). Human categorization. In N. Warren (ed.), *Studies in Cross-Cultural Psychology,* Vol. 1, 1–49. San Francisco: Academic Press

Sherry, D.F., and D.L. Schacter (1987). The evolution of multiple memory systems. *Psychological Review* 94: 439–54

Squire, L.R. (1987). *Memory and Brain.* New York: Oxford University Press

Streeter, L.A. (1976). Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature* 259: 39–41

Summerfield, Q. (1979). Use of visual information for phonetic perception. *Phonetica* 36: 314–31

Tulving, E. (1983). *Elements of Episodic Memory.* Oxford: Clarendon Press

———— (1985). Memory and consciousness. *Canadian Psychology* 26: 1–12

Volaitis, L.E., and J.L. Miller (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America* 92: 723–35

Walton, G.E., and T.G.R. Bower (1993). Amodal representation of speech in infants. *Infant Behavior and Development* 16: 233–43

Walton, G.E., A.G. Shoup-Pecenka and T.G.R. Bower (1991). Speech categorization in infants. *Journal of the Acoustical Society of America* 90: 2296

Werker, J. (1991). The ontogeny of speech perception. In I.G. Mattingly and M. Studdert-Kennedy (eds.), *Modularity and the Motor Theory of Speech Perception*, 91–109. Hillsdale, NJ: Erlbaum

Werker, J.F., and C.E. Lalonde (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology* 24: 672–83

Werker, J.F., and J.E.Pegg (1992). Infant speech perception and phonological acquisition. In C.A. Ferguson, L. Menn, and C. Stoel-Gammon (eds.), *Phonological Development: Models, Research, Implications*, 285–311. Timonium: York

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