4

The Development of Speech and Language

P.K. KUHL
Department of Speech and Hearing, University of Washington,
Seattle, WA 98195, U.S.A.

ABSTRACT

Infants learn language with remarkable speed. By the end of the second year they speak in sentences, sounding distinctly like a speaker of the "mother tongue." How does one individual acquire a specific language? Is it appropriate to call it "learning" in the traditional sense? Historically, two dramatically opposed views formed the cornerstones of the debate on language. According to one view, a universal grammar and phonology are innately provided and input serves to trigger this information. In the other view, no innate knowledge is provided and language is acquired through a process of feedback and reinforcement. Both theories are based on assumptions about the nature of language input to the child and the nature of the developmental change induced by input. New data reviewed here, showing the effects of early language experience on infants, suggest a theoretical revision. By one year of age, prior to the time infants begin to master higher levels of language, infants' perceptual and perceptual-motor systems have been altered by linguistic experience. Phonetic perception has changed dramatically to conform to the native-language pattern, and language-specific speech production has emerged. According to the model described here, this developmental change is caused by a complex "mapping" of linguistic input. This account is different in two respects from traditional views: (1) language input is not conceived of as triggering innately provided options, and (2) the kind of developmental change that occurs does not involve traditional Skinnerian learning, in which change is brought about through feedback and reinforcement. The consequences of this are described in a developmental theory at the phonetic level that holds promise for higher levels of language.

THEORIES OF LANGUAGE ACQUISITION AND THE "NATURE–NURTURE" DEBATE

Forty years ago, there was a historical confrontation between a strong nativist and a strong learning theorist. Chomsky's (1957) reply to Skinner's (1957) "Verbal Behavor" had just been published, reigniting the debate on the nature of language. In

'echanistic Relationships between Development and Learning

Chomsky's (1965, 1981) nativist view, universal rules encompassing the grammars and phonologies of all languages were innately specified. Language input served to "trigger" the appropriate subset of rules, and developmental change in language ability was viewed as biological growth akin to other bodily organs, rather than learning. In the Skinnerian view, language was explicitly learned. Language was brought about in the child through a process of feedback and reinforcement (Skinner 1957).

Both views made assumptions about three critical parameters: (a) the biological preparation that infants bring to the task of language learning, (b) the nature of language input, and (c) the nature of developmental change. Chomsky asserted, through the "poverty of the stimulus" argument, that language input to the child is greatly underspecified. Critical elements are missing; thus the necessity for innately specified information. Skinner viewed speech as simply another operant behavior, shaped through parental feedback and reinforcement like all other behaviors.

In the decades that have passed since these positions were developed, the debate has been played out at the syntactic, semantic, and phonological levels of language. In this chapter, I will primarily concentrate on the phonetic level, because it provides both a strong test of the opposing views and is amenable to detailed empirical manipulations.

Speech is readily accessible. One can study infants' perceptual abilities at birth to assess the initial state of their knowledge about speech perception, and examine at various points in development the results of nature's experiment — infants from various cultures who have been raised listening to vastly different languages. Adults from different cultures can be tested to assess the endstate of learning. Moreover, nonhuman animals can be tested to examine evolutionary constraints on sound perception and production. These studies have produced a great deal of data about the initial state of infant speech perception, the nature of developmental change that infants undergo in speech perception as they are exposed to a specific language, and the agent of developmental change, language input to the child.

The new data, demonstrating the effects of early language experience on infants, suggest a theoretical revision. By one year of age — prior to the time infants begin to master higher levels of language, such as sound-meaning correspondences, contrastive phonology, and grammatical rules — infants' perceptual and perceptual-motor systems have been altered by linguistic experience. Phonetic perception has changed dramatically to conform to the native-language pattern, and language-specific speech production has emerged. According to the model developed here, this developmental change is caused by a complex "mapping" of linguistic input. This account is different in two respects from traditional views: (1) language input is not conceived of as triggering innately provided options, nor as parameter setting, and (2) the kind of developmental change that occurs does not involve traditional Skinnerian learning, in which change is brought about through feedback and reinforcement.

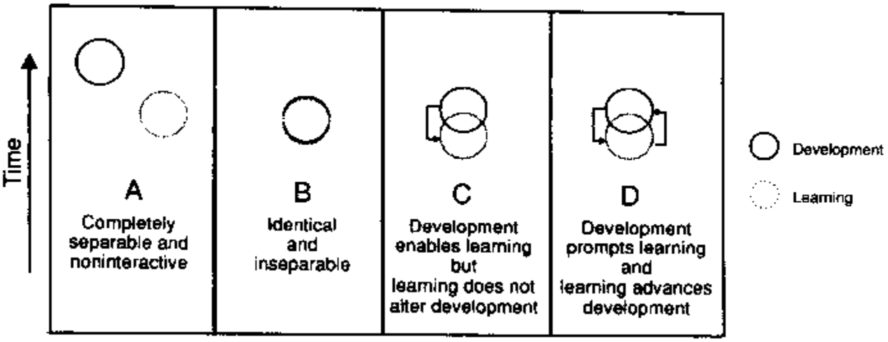
CONCEPTUAL DISTINCTIONS: "DEVELOPMENT" AND "LEARNING"

In human language, as well as in the development of species-typical behavior in animals, the relation between "development" (processes that cause change over time independent of experience) and "learning" (processes that depend on activity/experience of some kind) is a key to understanding the contribution of the organism and the environment. Are development and learning independent, and if not, how do they interact?

Four alternatives are conceptually illustrated in Figure 4.1. Development and learning could be thought of as completely separable processes (Figure 4.1A). Development follows a strict maturational course, and learning neither follows from nor leads to changes in the preestablished course of development. Alternatively, the two processes could be viewed as identical, so inseparable that we cannot pull them apart, even conceptually (Figure 4.1B).

More commonly, development and learning are thought of as separate and distinguishable processes that interact in one way or another (Figure 4.1C–D). Developmental psychologists, neuroscientists, and neurobiologists largely agree that interaction occurs between development and learning (see Bonhoeffer and Shatz, Doupe, Faneslow and Rudy, all this volume; Gopnik and Meltzoff 1997; Kuhl 1994). At issue, however, is exactly how the two systems interact, and particularly whether the interaction between development and learning is bidirectional.

Among the interactionist views, one model is that development enables learning, but that learning does not change the course of development, which unfolds more or less on its own timetable (Figure 4.1C). Learning is seen as capitalizing on the achievements of development and cannot occur unless a certain level of development has been achieved. The interaction is unidirectional, however. Development is not impacted by learning. In classical developmental psychology, this position is closest to the view of Piaget (1954). In modern neurobiology, the notion that there are constraints on learning, that development both prepares the organism and sets limits on learning, is consistent with this model (for the case of birdsong, see Marler 1974,



igure 4.1 Conceptual relations between development and learning.

and Doupe, this volume). Greenough and Black's (1992) "experience-expectant" plasticity, wherein changes in neural development are thought to precede and prepare an organism to react to a reliably present environmental stimulus, provides a detailed example of this model. In each of these cases, development is conceived of as both enabling and limiting learning, but learning does not alter the course of development.

There is an alternative interactionist view. This model describes development and learning as mutually affecting one another (Figure 4.1D). Development enables and even prompts learning, and learning in turn advances development. This view is closest to that developed by Vygotsky (1979). Vygotsky's theory, the "zone of proximal development" (ZPD), describes development at two levels: (1) the infant's actual developmental level, the level already achieved, and (2) the level that is just within reach. The ZPD is the difference between the two. In Vygotsky's view, environmental stimulation slightly in advance of current development (in the ZPD) results in learning and, when this occurrs, learning prompts development. A recent theory, proposed by developmental psychologists to account for a wide variety of cognitive and linguistic tasks, also provides a detailed model of mutual interaction between development and learning (Gopnik and Meltzoff 1997).

In linguistic theory, Chomsky's classic view that the growth of language is largely determined by a maturational process fits model C. Experience plays a role, but it is seen as triggering prespecified options or setting innately determined parameters rather than creating the form of language in any fundamental way. In contrast, the data reviewed here at the phonetic level of language come closer to the mutual interaction of model D. In the model of speech development I will describe, language input plays a significant role. Language input is mapped in a complex process that appears to code its subtle details. Input thus goes beyond simply triggering a prespecified option. Moreover, I argue that early mapping of the perceptual regularities of language allows infants to recognize words and phrases, thus advancing development.

In summary, there is a great deal of support for interactionist views (models C and D) over noninteractionist views (models A and B). While the relations between learning and development may differ across species and systems, there is an emerging consensus across diverse disciplines including neurobiology, psychology, linguistics, and neuroscience, that development and learning are not separate and distinct entities. The form of the interaction between the two remains a question, with a cutting-edge issue being whether (and how) learning can alter development. The model I will propose here is based on recent research of speech development and goes some distance toward addressing this issue.

THE FOCUS OF INQUIRY: DEVELOPMENTAL CHANGE IN PHONETIC PERCEPTION

One of the puzzles in language development is to explain the orderly transitions that all infants go through during development. Infants the world over achieve certain milestones in linguistic development at roughly the same time, regardless of the language they are exposed to. Moreover, developmental change can also include cases in which infants' early skills exceed their later ones. Explaining these transitions is one of the major goals of developmental linguistic theory.

One of these transitions occurs in *speech perception*. At birth, infants discern differences between all the phonetic units used in the world's languages (Eimas et al. 1987). All infants show these universal skills, regardless of the language environment in which they are being raised. Data on nonhuman animals' perception of speech suggest that the ability to partition the basic building blocks of speech is one delivered by evolution (Kuhl 1991b).

When do infants from different cultures begin to diverge? Infants' initial "language-general" abilities are highly constrained just one year later. By the end of the first year, infants fail to discriminate foreign-language contrasts they once discriminated (Werker and Tees 1984), resembling the adult pattern. Adults often find it difficult to perceive differences between sounds not used to distinguish words in their native language. Adult native speakers of Japanese, for example, have great difficulty discriminating American English /r/ and /l/ (Strange 1995; Best 1993), and American English listeners have great difficulty hearing the difference between Spanish /b/ and p/ (Abramson and Lisker 1970). Infants' abilities change over a 6-month period. A recent study we have just completed in Japan shows, for example, that at 6 months of age Japanese infants respond to the /r – l/ distinction and are as accurate in perceiving it as American 6-month-old infants. By 12 months, Japanese infants no onger demonstrate this ability, even though American infants at that same age have become even better at discriminating the two sounds (Kuhl, Kiritani, Deguchi, Hayashi, Stevens and Morton, in preparation).

A similar transition occurs in *speech production*. Regardless of culture, all infants progress through a set of universal stages during the first year (Ferguson et al. 1992). By the end of the first year, however, the utterances of infants reared in different ountries begin to diverge, reflecting the ambient language (de Boysson-Bardies 993). In adulthood, the speech motor patterns that we initially learned contribute o our "accents" when attempting to speak another language.

These transitions in speech present one of the most intriguing problems in anguage acquisition. What causes these changes? The thesis developed here is that nguistic experience produces a special kind of developmental change. Language nput alters the brain's processing of speech, resulting in the creation of complex mental maps for speech. The mapping "warps" underlying dimensions, altering erception in such a way as to highlight native-language categories while making oreign-language categories less discriminable. This mapping is not like traditional earning. It does not depend on external reinforcement and appears to be unconscious nd long-lasting.

LANGUAGE EXPERIENCE ALTERS PERCEPTION

Language experience produces a mapping that alters phonetic perception. A research finding that helps explain how this occurs is called the "perceptual magnet effect." It is observed when tokens perceived as exceptionally good representatives of a phonetic category ("prototypes") are used in tests of speech perception (Kuhl 1991a). Our results show that phonetic prototypes function like *perceptual magnets* for other sounds in the category. When listeners hear a phonetic prototype and attempt to discriminate it from sounds that surround it in acoustic space, the prototype displays an attractor effect on the surrounding sounds (Figure 4.2). It perceptually pulls other members of the category toward it, making it difficult to hear differences between the prototype and surrounding stimuli. Poor instances from the category (nonprototypes) do not function in this way. A variety of experimental tasks produce this result with both consonants and vowels (Iverson and Kuhl 1995, 1996; Sussman and Lauckner-Morano 1995). Other studies confirm listeners' skills in identifying phonetic prototypes and show that they are language specific (Kuhl 1992; Miller 1994; Willerman and Kuhl 1996).

Developmental tests revealed that the perceptual magnet effect was exhibited by 6-month-old infants for the sounds of their native language (Kuhl 1991a). Moreover, cross-language experiments demonstrated that the magnet effect is the product of linguistic experience (Kuhl et al. 1992). In the cross-language experiment, infants in the United States and Sweden were tested. The infants from both countries were tested with two vowel prototypes, an American English vowel prototype, /i/ (as in "peep"), and a Swedish vowel prototype, /y/ (as in "fye"). The results demonstrated that the perceptual magnet effect in 6-month-old infants was affected by exposure to a particular language. American infants demonstrated the magnet effect only for the American English /i/; they treated the Swedish /y/ like a nonprototype. Swedish

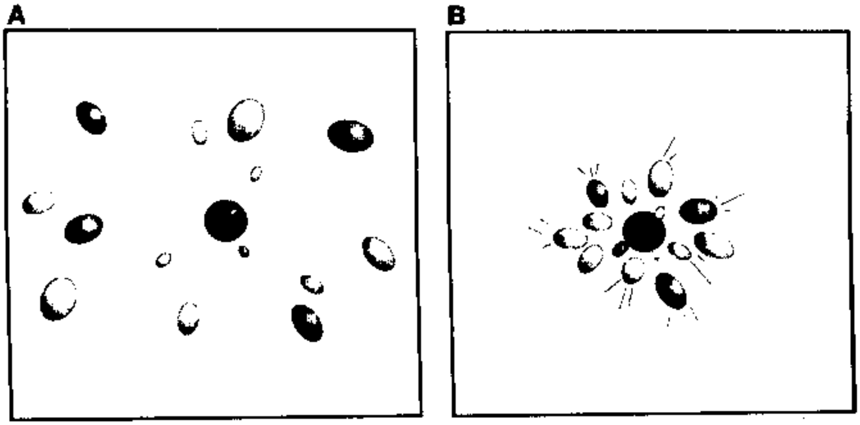


Figure 4.2 The *perceptual magnet effect.* When a variety of sounds in a category surround the category prototype (A), they are perceptually drawn toward the prototype (B). The prototype appears to function like a magnet for other stimuli in the category.

infants showed the opposite pattern, demonstrating the magnet effect for the Swedish /y/ and treating the American English /i/ as a nonprototype. This is the youngest age at which language experience has been shown to affect phonetic perception.

The perceptual magnet effect thus occurs prior to word learning. What this means is that in the absence of formal language understanding or use — before infants utter or understand their first words — infants' perceptual and perceptual-motor systems strongly conform to the characteristics of the ambient language. We previously believed that word learning caused infants to recognize that phonetic changes that they could hear, such as the change that Japanese infants perceived between /r/ and /l/, did not change the meaning of a word in their language. This discovery was thought to cause the change in phonetic perception. We now know that just the opposite is true. Language input sculpts the brain to create a perceptual system that highlights the contrasts used in the language, while deemphasizing those that do not, and this happens prior to word learning. The change in phonetic perception thus assists word learning, rather than the reverse.

Further tests on adults suggested that the magnet effect distorted perception to highlight sound contrasts in the native language. Studies on the perception of the phonetic units /r/ and /l/ as in the words "rake" and "lake," illustrate this point. The /r – l/ distinction is one notoriously difficult for Japanese speakers, and our studies sought to determine how adults from different cultures perceived these two sounds. To conduct the study, we used computer-synthesized syllables beginning with /r/ and /l/, spacing them at equal physical intervals in a 2-dimensional acoustic grid (Iverson and Kuhl 1996) (Figure 4.3A). American listeners identified each syllable as beginning with either /r/ or /l/, rated its category goodness, and estimated the perceived similarity for all possible pairs of stimuli using a scale from "1" (very dissimilar) to "7" (very similar). Similarity ratings were scaled using multidimensional scaling (MDS) techniques. The results revealed that perception distorts physical space. The physical (acoustic) differences between pairs of stimuli were equal; however, perceived distance was "warped" (Figure 4.3B). The perceptual space around the best

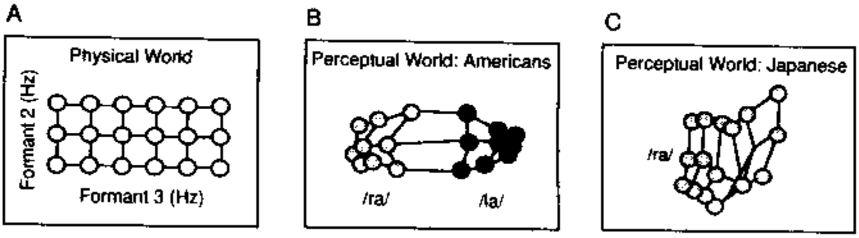


Figure 4.3 Physical (acoustic) versus perceptual distance. Consonant tokens of /r/ and /l/ were generated to be equally distant from one another in acoustic space (A). However, American listeners perceive perceptual space as "shrunk" near the best instances of /r/ (gray dots) and /l/ (black dots) and "stretched" at the boundary between the two (B). Japanese listeners' perceptual world differs dramatically; neither magnet effects nor a boundary between the two categories are seen (C).

/r/ and the best /l/ was greatly reduced, as predicted by the perceptual magnet effect, while the space near the boundary between the two categories was expanded.

This experiment has now been done using Japanese monolingual listeners (Iverson, Kuhl, Yamada, Tohkura, and Stevens, in preparation) and the results show a strong contrast in the way the /r – l/ stimuli are perceived by American and Japanese adults (Figure 4.3C). Japanese adults hear almost all the sounds as /r/; there is no /l/ in Japanese. More striking is the complete absence of magnet and boundary effects in the Japanese MDS solution. The results suggested that linguistic experience results in the formation of *perceptual maps* specifying the perceived distances between stimuli. These maps increase internal category cohesion while maximizing the distinction between categories. The critical point for theory is that neither group perceives physical reality, the actual physical differences between sounds. For each language group, experience has altered perception to create a language-specific map of auditory similarities and differences, one that highlights the sound contrasts of the speaker's native language. These mental maps for speech are the front-end of the language mechanism. They filter language in a way that promotes semantic and syntactic analysis.

The theoretical position developed here is that the mental maps for speech are developed early in infancy, prior to the development of word acquisition. The mapping of phonetic information is seen as enabling infants to recognize word candidates. For example, our work shows that Japanese infants fail to discriminate American English /r/ from /l/ at 12 months of age, though they did so perfectly well at 6 months of age (Kuhl et al., in preparation). This is argued to assist Japanese infants in word recognition. The collapsing of /r/ and /l/ into a single category makes it possible for Japanese infants to perceive their parents' productions of /r/-like and /l/-like sounds as one entity at 12 months, when the process of word acquisition begins. If they did not do so, it would presumably make it more difficult to map sound patterns on to objects and events.

The view that phonetic mapping supports the recognition of higher-order units is supported by data showing that slightly later in development infants use information about phonetic units to recognize word-like forms. Work by Jusczyk and his colleagues shows that by 9 months of age, infants prefer word patterns that are typical of the native language, which requires recognition of native-language phonetic units (Jusczyk et al. 1993). Infants have also been shown capable of learning the statistical probabilities of sound combinations contained in artificial words (Saffran et al. 1996). Infants' mapping at the phonetic level is thus seen as assisting infants in "chunking"the sound stream into higher-order units.

These studies indicate that during the first year of life, infants come to recognize the perceptual properties of their native language. To do this, infants must be mentally storing those properties in some form. This occurs in the absence of any formal instruction or reinforcement of the infant's behavior. In this sense, the "learning" that transpires is outside the realm of the historical versions of learning described by psychologists.

A THEORY OF SPEECH DEVELOPMENT

These findings have been incorporated in a three-step theory of speech development, called the *Native Language Magnet* (NLM) model (Kuhl 1994). NLM describes infants' initial state as well as changes brought about by experience with language (Figure 4.4). The model demonstrates how infants' developing native-language speech representations might alter both speech perception and production. The example developed here is for vowels, though the same principles apply to consonant perception.

Phase 1 describes infants' initial abilities. At birth, infants partition the sound stream into gross categories separated by natural auditory boundaries (Figure 4.4A). As shown, perceptual boundaries divide a hypothetical vowel space, separating the vowels of all languages. According to NLM, infants' abilities at this stage do not depend on specific language experience. The boundaries initially structure perception in a phonetically relevant way. However, they are not due to a "language module." This notion is buttressed by the fact that these same perceptual boundary

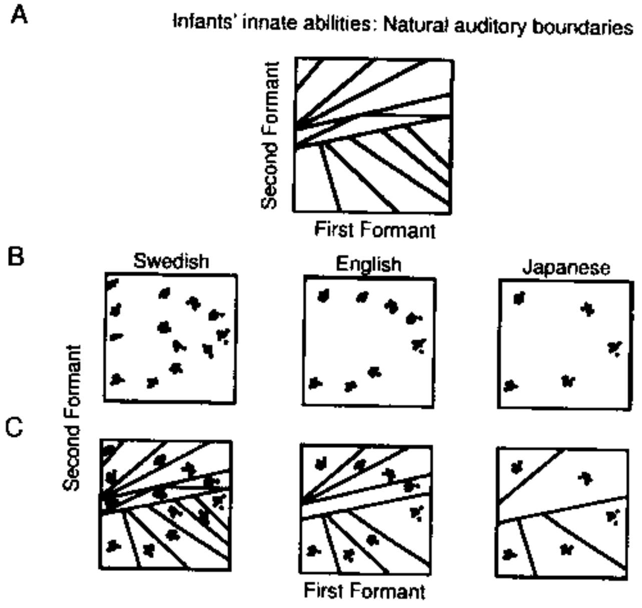


Figure 4.4 The Native Language Magnet (NLM) model. (A) At birth, infants perceptually partition the acoustic space underlying phonetic distinctions in a language-universal way. They are capable of discriminating all phonetically relevant differences in the world's anguages. (B) By 6 months of age, infants reared in different linguistic environments show an effect of language experience. Infants store incoming vowel information in memory in ome form. The resulting representations (shown by the dots) are language specific and reflect he distributional properties of vowels in the three different languages. (C) After language-pecific magnet effects appear, some of the natural boundaries that existed at birth "disappear." nfants now fail to discriminate foreign-language contrasts they once discriminated.

phenomena are exhibited in the same places in acoustic space by nonhuman animals (Kuhl 1991b).

Phase 2 describes the vowel space at 6 months of age for infants reared in three very different language environments: Swedish, English, and Japanese (Figure 4.4B). By 6 months of age, infants show more than the innate boundaries shown in Phase 1. By 6 months, our calculations indicate that infants have heard hundreds of thousands of instances of particular vowels. According to NLM, infants represent this information in memory in some form. The distributional properties of vowels heard by infants being raised in Sweden, America, and Japan differ. As shown in 4B, their stored representations also differ, reflecting these distributional differences. In each case, linguistic experience has produced stored representations that reflect the vowel system of the ambient language. Language-specific magnet effects, produced by the stored representations, are exhibited by infants at this stage.

Phase 3 shows how magnet effects recursively alter the initial state of speech perception. Magnet effects cause certain perceptual distinctions to be minimized (those near the magnet attractors) while others are maximized (those near the boundaries between two magnets). The consequence is that some of the boundaries that initially divided the space "disappear" as the perceptual space is reconfigured to incorporate a language's particular magnet placement (Figure 4.4C). Sensory perception has not changed; higher-order memory and representational systems have altered infants' abilities. Magnet effects functionally erase certain boundaries — those relevant to foreign but not native languages. At this stage, a perceptual space once characterized by simple boundaries has been replaced by a warped space dominated by magnets.

Infants at 6 months of age have no awareness of the fact that sound units are used contrastively in language to name things. Yet the infant's perceptual system organizes itself to reflect language-specific phonetic categories. At the next stage in linguistic development, when infants acquire word meanings by relating sounds to objects and events in the world, the language-specific mapping that has already occurred should greatly assist this process.

NLM theory offers an explanation for the transition in speech perception. A developing magnet pulls sounds that were once discriminable towards it, making them less discriminable. Magnet effects should therefore developmentally precede changes in infants' perception of foreign-language contrasts; preliminary data indicate that they do (Werker and Polka 1993). The magnet effect also helps account for the results of studies on the perception of sounds from a foreign language by adults (Best 1993; Flege 1993). For example, NLM theory may help explain Japanese listeners' difficulty with American /r/ and /l/. The magnet effect for the Japanese /r/ category prototype (which is neither American /r/ nor /l/) will attract both /r/ and /l/, making the two sounds difficult for native-speaking Japanese people to discriminate. NLM theory argues that early experience established perceptual "filters" through which language passes. In this view, one's primary language will affect how other languages are perceived.

REINTERPRETING "CRITICAL PERIODS"

The traditional literature on critical periods views them as a strictly timed developmental process, independent of learning and other factors. The critical period thus defines a "window of opportunity" during which environmental stimulation is effective in producing developmental change. Once the window closes, environmental stimulation is no longer effective. Current studies show that the window can be stretched by a variety of factors (Doupe, this volume); however, the notion that the critical period as a process governed by maturation remains.

There is an alternative possibility. Later learning may be limited by an *interference* factor. For example, if learning "commits" neural structure in some way (e.g., NLM's argument that learning involves the creation of mental maps for speech), future learning is expected to be affected by this commitment. The mechanisms governing an organism's general ability to learn may not have changed. Rather, initial learning may result in a structure that reflects environmental input and, once committed, the learned structure may interfere with the processing of information that does not conform to the learned pattern. On this account, initial learning can alter future learning independent of a strictly timed period.

The interference view may account for some aspects of second-language learning. When acquiring a second language, certain phonetic distinctions are notoriously difficult to master, both in speech perception and production. Take the case of the /r – l/ distinction for native speakers of Japanese. Hearing the distinction and producing it is very difficult for native speakers of Japanese. According to NLM, this is the case because exposure to Japanese early in life altered the Japanese infant's perceptual system, resulting in magnet effects for the Japanese phoneme /r/, but not for American English /l/ (which is not phonemic in Japanese). Once in place, the magnet effects appropriate for Japanese would not make it easy to process American English. Both American English /r/ and /l/ would be assimilated to Japanese /r/. Thus, interference effects could make it much more difficult to acquire new phonetic categories later in life.

On this "interference" account, plasticity would be governed from a statistical standpoint. When additional input does not cause the overall statistical distribution to change substantially, the organism becomes less sensitive to input. Hypothetically, for instance, the infants' representation of the vowel /a/ might not change when the one millionth token of the vowel /a/ is heard. Plasticity might thus be independent of time, but dependent on the amount and variability provided by experience. At some time in the life of the organism one could conceive of a point where new input no longer alters the underlying distribution, and this could, at least in principle, define the end of the "critical period" for learning.

Early in life, interference effects are minimal and new categories can be acquired because input continues to revise the statistical distribution. In this context it is interesting to note that anecdotal evidence suggests that infants exposed to two languages do much better if each parent each speaks one of the two languages, rather both parents speaking both languages. This may be the case because it is easier to map two different sets of phonetic categories (one for each of the two languages) if there was some way to keep them perceptually separate. Males and females produce speech in different frequency ranges, and this could make it easier to maintain separation.

These two factors — a maturationally defined temporal window and initial learning that makes later learning more difficult — could both be operating to produce constraints on learning a second language later in life. If a maturational process induces "readiness" at a particular time, input that misses this timing could reduce learning. At the same time, an "interference" factor might provide an independent mechanism that contributes to the difficulty in readily learning a second language in adulthood.

THE PERCEPTUAL-MOTOR LINK: SPEECH PRODUCTION

Infants not only learn the perceptual properties of their native language, but become proficient speakers of the language. Once learned, speaking patterns become difficult to alter. Speakers who learn a second language later in life, for example, produce it with an "accent" typical of their primary language. Most speakers of a second language would like to speak like a native speaker, without an accent, but this is difficult to do.

When do we adopt the speech patterns that will mark us as native speakers of a particular language? Developmental studies suggest that by 1 year of age language-specific patterns of speech production appear in infants' spontaneous utterances (de Boysson-Bardies 1993; Vihman and de Boysson-Bardies 1994). However, the fundamental capacity to reproduce the sound patterns one hears is in place much earlier. In a recent study, Kuhl and Meltzoff (1996) recorded infant utterances at 12, 16, and 20 weeks of age while the infants watched and listened to a video recording of a woman producing a vowel, either /a/, /i/, or /u/. Infants watched the video for 5 minutes on each of three consecutive days. Infants' utterances were analyzed both perceptually (phonetic transcription) and instrumentally (computerized spectrographic analysis).

The results showed that there was developmental change in infants' vowel productions between 12 and 20 weeks of age. The areas of vowel space occupied by infants' /a/, /i/, and /u/ vowels become progressively more tightly clustered at each age, and by 20 weeks, a "vowel triangle" typical of that produced in every language of the world, had emerged in infants' own region of the vowel space (Figure 4.5). This suggested the possibility that infants were listening to language and attempting to imitate vocally the sound patterns they heard (Kuhl and Meltzoff 1996).

Direct evidence that infants were vocally imitating was also obtained in the study. By 20 weeks, infants were shown to reproduce the vowels they heard. Infants exposed to /a/ were more likely to produce /a/ than when exposed to either /i/ or /u/; similarly, infants exposed to either /i/ or /u/ were more likely to produce the vowel in that

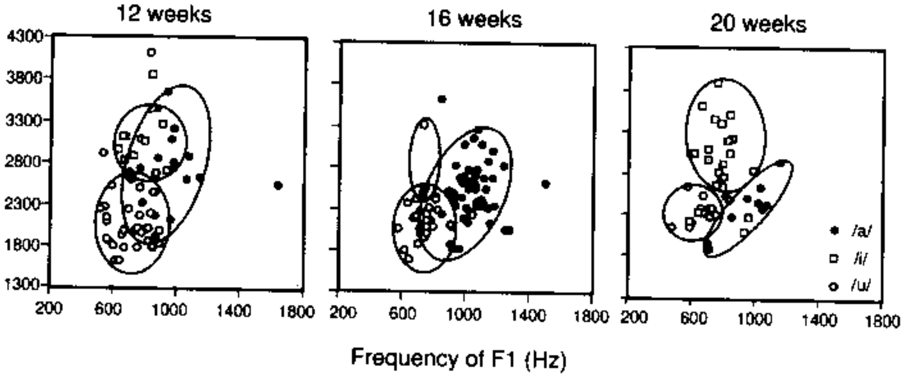


Figure 4.5 The location of /a/, /i/, and /u/ vowels produced by 12-, 16-, and 20-week-old infants. Infants' vowel productions show progressively tighter clustering in vowel space over the 8-week period and reflect differences between the three vowel categories seen in adults' productions.

condition than when listening to either of the two alternate vowels. The total amount of exposure to a specific vowel in the laboratory was only 15 minutes, yet this was sufficient to influence infants' productions. If 15 minutes of laboratory exposure to a vowel is sufficient to influence infants' vocalizations, then listening to ambient language for weeks would be expected to provide a powerful influence on infants' production of speech. These data suggest that infants' stored representations of speech not only alter infant perception, but alter production as well, serving as auditory patterns that guide motor production. Stored representations are thus viewed as the common cause for both the tighter clustering observed in infant vowel production and the tighter clustering observed in infant vowel perception (Figure 4.6).

This pattern of learning and self-organization, in which perceptual patterns stored in memory serve as guides for production, is strikingly similar to that seen in other domains involving auditory-perceptual learning, such as birdsong (Doupe, this volume), and in visual-motor learning, such as gestural imitation (Meltzoff and Moore 1977, 1994). In each of these cases, perceptual experience establishes a representation that guides sensory-motor learning. In the case of infants and speech, perception affects production in the earliest stages of language learning, reinforcing the idea that the speech motor patterns of a specific language are are formed very early in life. This would also help explain the fact that speech motor patterns are difficult to alter when we attempt to learn a second language.

POLYMODAL SPEECH REPRESENTATION

The link between perception and production can be seen in another experimental situation. Speech perception in adults is strongly affected by the sight of a talker's

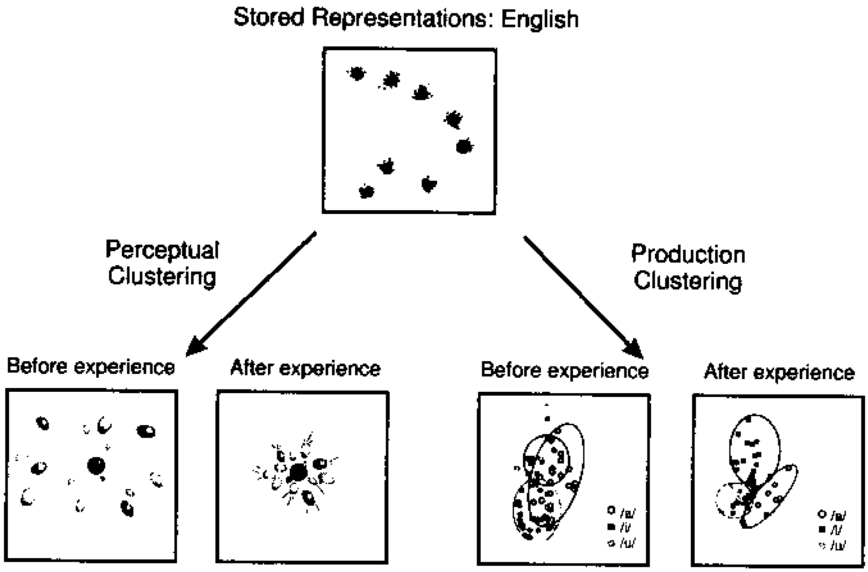


Figure 4.6 Stored representations of native-language speech affect both speech perception, producing the perceptual clustering evidenced by the magnet effect, as well as speech production, producing the increased clustering seen in infants' vocalizations over time.

mouth movements during speech, indicating that our representational codes for speech contain not both auditory and visual information. One of the most compelling examples of the polymodal nature of speech is auditory-visual illusions that result when discrepant information is sent to two separate modalities. One such illusion occurs when auditory information for /b/ is combined with visual information for /g/ (McGurk and MacDonald 1976; Green et al. 1991; Kuhl et al. 1994; Massaro 1987). Perceivers report the phenomenal impression of an intermediate articulation (/da/ or /tha/) despite the fact that this information was not delivered to either sense modality. This is a very robust phenomenon and is readily obtained even when the information from the two modalities comes from different speakers, such as when a male voice is combined with a female face (Green et al. 1991). In this case, there is no doubt that the auditory and visual signals do not belong together. Yet the illusion is still unavoidable — our perceptual systems combine the multimodal information (auditory and visual) to give a unified percept.

Language experience affects this auditory-visual illusion. When native speakers watch and listen to incongruent audiovisual speech signals pronounced by a foreign speaker, they show increased auditory-visual effects — greater numbers of illusory responses occur (Kuhl et al. 1994; Sekiyama and Tohkura 1993). The auditory information in foreign speech does not precisely match the stored representations of native-language speech. When this occurs, visual information may be relied upon to a greater extent, reinforcing the idea that speech representations are polymodally mapped.

Young infants also appear to represent speech polymodally. Infants just 18–20 weeks old recognize auditory-visual correspondences for speech, akin to what we as adults do when we lipread; in these studies, infants looked longer at a face pronouncing a vowel that matched the vowel sound they heard rather than a mismatched face (Kuhl and Meltzoff 1982). Young infants demonstrate knowledge about both the auditory and visual information contained in speech, supporting the notion that infants' stored speech representations contain information of both kinds.

NATURE OF LANGUAGE INPUT TO THE CHILD

The studies just reviewed attest to the impact of language input on infants. If language input produces change in both speech perception and production, we need to know a great deal more about it. How much and what kind of speech do infants hear?

Estimates indicate that a typical listening day for a 2-year-old includes 20,000 to 40,000 words (Chapman et al. 1992). Speech addressed to infants (often called "motherese" or "parentese") is unique: it has a characteristic prosodic structure that includes a higher pitch, a slower tempo, and exaggerated intonation contours, and it is syntactically and semantically simplified. Research supports the idea that this speaking style is near universal in the speech of caretakers around the world and that infants prefer it (Fernald 1985; Fernald and Kuhl 1987). The motherese pattern of speech, with its higher pitch and expanded intonation contours, is attractive to infants but may not be necessary for infant learning.

In new studies, we have uncovered another modification made by parents when addressing infants that may be much more important to infant learning. We examined natural language input at the phonetic level to infants in the United States, Russia, and Sweden (Kuhl et al. 1997). The study shows that across three very diverse languages, infant-directed speech exhibited a universal alteration of phonetic units when compared to adult-directed speech. Parents addressing their infants produced acoustically more extreme tokens of vowel sounds, resulting in a "stretching" of the acoustic space encompassing the vowel triangle (Figure 4.7). A stretched vowel triangle not only makes speech more discriminable for infants, it highlights critical spectral parameters that allows speech to be produced by the child. The results suggest that at the phonetic level of language, linguistic input to infants provides exceptionally well-specified information about the units that form the building blocks for words.

A stretched vowel space is not necessary from the standpoint of the infant's capacity to distinguish vowels. The formant frequency changes from adult-directed to infant-directed speech were substantial, ones that would clearly be registered by the infant auditory system. However, previous data on infants' capacities to discern subtle differences between vowels indicate that infants are capable of hearing

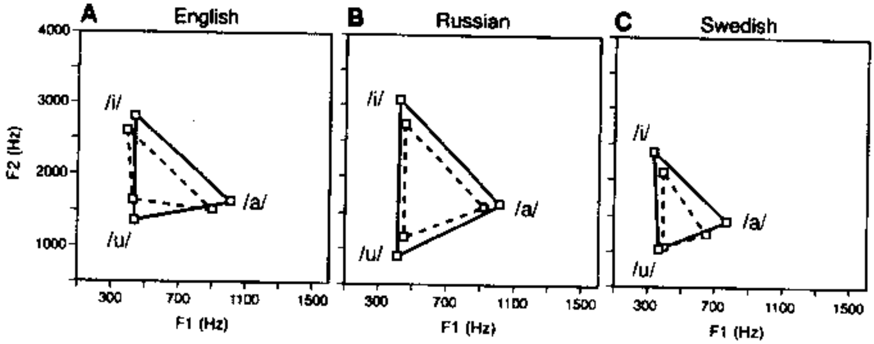


Figure 4.7 The vowel triangle of maternal speech directed toward infants (solid line) across three diverse languages shows a "stretching" relative to the adult-directed vowel triangle (dashed line), an effect that both makes vowels more discriminable and highlights the abstract features that infants must use to produce speech themselves.

differences a great deal smaller than those produced by mothers in the study (Kuhl 1991b).

If not required for infant discrimination, what function does a stretched vowel space serve? We hypothesized that stretching the vowel triangle could benefit infants in three ways. First, it increases the distance between vowels, making them more distinct from one another. In recent studies, language-delayed children showed substantial improvements in measures of speech and language after treatment in a program in which they listened to speech altered by computer to exaggerate phonetic differences (Merzenich et al. 1996; Studdert-Kennedy and Mody 1995; Tallal et al. 1996). Normally developing infants may benefit similarly from the enhanced acoustic differences provided in infant-directed speech.

Second, to achieve the stretching, mothers produce vowels that go beyond those produced in typical adult conversation. From both an acoustic and articulatory perspective, these vowels are "hyperarticulated" (Lindblom 1990). Hyperarticulated vowels are perceived by adults as "better instances" of vowel categories (Iverson and Kuhl 1995; Johnson et al. 1993), and laboratory tests show that when listening to good instances of phonetic categories, infants show greater phonetic categorization ability.[1] The present study shows that hyperarticulated vowels are a part of infants'

---

[1] The American English "prototype /i/ vowel used in Kuhl's (1991b; Kuhl et al. 1992; Iverson and Kuhl 1995) laboratory studies was the male speakers' average /i/ in Peterson and Barney's (1952) data set. If the formant values for Peterson and Barney's female average /i/ are compared with mothers' productions of /i/ in Kuhl et al. (1997), the formant values fall very near the mean of American English mothers' infant-addressed /i/. Thus, the earlier findings showing infants' superior categorization when listening to a "prototype" and the recent results showing that mothers hyperarticulate vowels when addressing their infants are consistent.

linguistic experience and raises the possibility that they may play an important role in the development of infants' vowel categories.

Third, expanding the vowel triangle allows mothers to produce a greater variety of instances representing each vowel category without creating acoustic overlap between vowel categories. Greater variety may cause infants to attend to nonfrequency-specific spectral dimensions that characterize a vowel category, rather than on any one particular set of frequencies that the mother uses to produce a vowel (Lively et al. 1993). Converting the formant values to spectral features shows that infant-directed speech maximizes the featural contrast between vowels (Kuhl et al. 1997). This is especially critical for infants because they cannot duplicate the absolute frequencies of adult speech — their vocal tracts are too small (Kent and Murray 1982). In order to speak, infants must reproduce the appropriate spectral dimensions in their own frequency range (Kuhl and Meltzoff 1996). Our recent study indicates that maternal language input emphasizes these dimensions.

NEURAL CORRELATES OF SPEECH PROCESSING

Various techniques of neuroscience (PET, MRI, MRI, and MEG) have been applied to phonetic processing in adults (Näätänen et al. 1997), and infant studies are beginning to appear. High-density event-related potentials (ERPs) have recently been used to study word processing in young children (Mills et al. 1993). These techniques are now being applied to speech processing in infants (Dehaene-Lambertz and Dehaene 1994) and newborns (Cheour-Luhtanen et al. 1995).

Additional work in adults suggests that the mismatched negativity (MMN) response, an ERP component thought to reflect preattentive auditory processes (Sharma et al. 1993), and magnetoencephalography (MEG) will provide interesting measures of language-specific speech perception. Recent studies suggest that phonetic prototypes provide a particularly sensitive measure of language experience (Näätänen et al. 1997). Such measures hold promise for mapping the brain's responses to speech over the lifespan.

CONCLUSIONS

In the first year of life, infants learn a great deal about the perceptual characteristics of their native language. Learning subsequently alters the perception and production of speech. According to Kuhl's "Native Language Magnet" model, perceptual learning early in life results in the formation of stored representations that capture native-language regularities. The theory emphasizes the role of linguistic input. Input does not act like a trigger for innately stored information. Rather, it is mapped in such a way as to "warp" the underlying acoustic space. Stored representations act like *perceptual magnets* for similar patterns of sound, resulting in maps that specify perceived distances between sounds. The map shrinks perceptual distances near a category's most typical instances and stretches perceptual distances between categories. Perceptual maps differ in adults who speak different languages. The magnet effects and the perceptual maps they produce also affect speech production. This helps explain speech development in infants and helps explain why, as adults, we do not hear or produce foreign-language sounds very well. Current work is aimed at examining the brain changes that accompany language learning using the techniques of modern neuroscience.

ACKNOWLEDGMENT

Supported by grants to the author from the National Institutes of Health (DC 00520, HD 18286, and HD 22514).

REFERENCES

Abramson, A.S., and L. Lisker. 1970. Discriminability along the voicing continuum: Cross-language tests. In: Proc. of the 6th Intl. Congress of Phonetic Sciences, Prague 1967, pp. 569–573. Prague: Academia.

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

Chapman, R.S., N.W. Streim, E.R. Crais, D. Salmon, E.A. Strand, and N.A. Negri. 1992. Child talk: Assumptions of a developmental process model for early language learning. In: Processes in Language Acquisition and Disorders, ed. R.A. Chapman, pp. 3–19. St. Louis: Mosby Year Book.

Cheour-Luhtanen, M., K. Alho, T. Kujala, K. Sainio, K. Reinikainen, M. Renlund, O. Aaltonen, O. Eerola, and R. Näätänen. 1995. Mismatch negativity indicates vowel discrimination in newborns. *Hear. Res.* **82**:53–58.

Chomsky, N. 1957. A review of B.F. Skinner's Verbal Behavior. *Language* **35**:26–58.

Chomsky, N. 1965. Aspects of the theory of syntax. Cambridge: MIT Press.

Chomsky, N. 1981. Rules and Representations. New York: Columbia Univ. Press.

de Boysson-Bardies, B. 1993. Ontogeny of language-specific syllabic productions. In: Developmental Neurocognition: Speech and Face Processing in the First Year of Life, ed. B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, and J. Morton, pp. 353–363. Dordrecht: Kluwer Academic.

Dehaene-Lambertz, G., and S. Dehaene. 1994. Speed and cerebral correlates of syllable discrimination in infants. *Nature* **370**:292–295.

Eimas, P.D., J.L. Miller, and P.W. Jusczyk. 1987. On infant speech perception and the acquisition of language. In: Categorical Perception: The Groundwork of Cognition, ed. S. Harnad, pp. 161–195. New York: Cambridge Univ. Press.

Ferguson, C.A., L. Menn, and C. Stoel-Gammon, eds. 1992. Phonological Development: Models, Research, Implications. Timonium, MD: York.

Fernald, A. 1985. Four-month-old infants prefer to listen to motherese. *Infant Behav. Dev.* **8**:181–195.

Fernald, A., and P. Kuhl. 1987. Acoustic determinants of infant preference for motherese speech. *Infant Behav. Dev.* **10**:279–293.

Flege, J.E. 1993. Production and perception of a novel, second-language phonetic contrast. *J. Acoust. Soc. Am.* **93**:1589–1608.

Gopnik, A., and A.N. Meltzoff. 1997. Words, Thoughts, and Theories. Cambridge: MIT Press.

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. *Percept. Psychophys.* **50**:524–536.

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

Hillenbrand, J., L.A. Getty, and M.J. Clark. 1995. Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* **97**:3099–3111.

Iverson, P., and P.K. Kuhl. 1995. Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *J. Acoust. Soc. Am.* **97**:553–562.

Iverson, P., and P.K. Kuhl. 1996. Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *J. Acoust. Soc. Am.* **99**:1130–1140.

Johnson, K., E. Flemming, and R. Wright. 1993. The hyperspace effect: Phonetic targets are hyperarticulated. *Language* **69**:505–528.

Jusczyk, P.W., A.D. Friederici, J.M.I. Wessels, V.Y. Svenkerud, and A.M. Jusczyk. 1993. Infants' sensitivity to the sound patterns of native language words. *J. Mem. Lang.* **32**:402–420.

Kent, R.D., and A.D. Murray. 1982. Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *J. Acoust. Soc. Am.* **72**:353–365.

Kuhl, P.K. 1991a. Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Percept. Psychophys.* **50**:93–107.

Kuhl, P.K. 1991b. Perception, cognition, and the ontogenetic and phylogenetic emergence of human speech. In: Plasticity of Development, ed. S.E. Brauth, W.S. Hall, and R.J. Dooling, pp. 73–106. Cambridge: MIT Press.

Kuhl, P.K. 1992. Infants' perception and representation of speech: Development of a new theory. In: Proc. of the Intl. Conf. on Spoken Language Processing, ed. J.J. Ohala, T.M. Nearey, B.L. Derwing, M.M. Hodge, and G.E. Wiebe, pp. 449–456. Edmonton: Univ. of Alberta.

Kuhl, P.K. 1994. Learning and representation in speech and language. *Curr. Opin. Neurobiol.* **4**:812–822.

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

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

Kuhl, P.K., and A.N. Meltzoff. 1996. Infant vocalizations in response to speech: Vocal imitation and developmental change. *J. Acoust. Soc. Am.* **100**:2425–2438.

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: Proc. of the Intl. Conf. on Spoken Language Processing, pp. 539–542. 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**:606–608.

Lindblom, B. 1990. Explaining phonetic variation: A sketch of the HandH theory. In: Speech Production and Speech Modeling, ed. W.J. Hardcastle and A. Marchal, pp. 403–439. Dordrecht: Kluwer Academic.

Lively, S.E., J.S. Logan, and D.B. Pisoni. 1993. Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *J. Acoust. Soc. Am.* **94**:1242–1255.

Marler, P. 1974. Constraints on learning: Development of bird song. In: Ethology and Psychiatry, ed. W.F. Norman, pp. 69–83. Toronto: Univ. Toronto Press.

Massaro, D.W. 1987. 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**:746–748.

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

Meltzoff, A.N., and M.K. Moore. 1994. Imitation, memory, and the representation of persons. *Infant Behav. Dev.* **17**:83–99.

Merzenich, M.M., W.M. Jenkins, P. Johnston, C. Schreiner, S.L. Miller, and P. Tallal. 1996. Temporal processing deficits of language-learning impaired children ameliorated by training. *Science* **271**:77–81.

Miller, J.L. 1994. On the internal structure of phonetic categories: A progress report. *Cognition* **50**:271–285.

Mills, D.L., S.A. Coffey-Corina, and H.J. Neville. 1993. Language acquisition and cerebral specialization in 20-month-old infants. *J. Cogn. Neurosci.* **5**:317–334.

Näätänen, R., A. Lehtokoski, M. Lennes, M. Cheour, M. Huotilainen, A. Iivonen, M. Vainio, P. Alku, R.J. Ilmoniemi, A. Luuk, J. Allik, J. Sinkkonen, and K. Alho. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* **385**: 432–434.

Peterson, G.E., and H.L. Barney. 1952. Control methods used in a study of the vowels. *J. Acoust. Soc. Am.* **24**:175–184.

Piaget, J. 1954. The Construction of Reality in the Child. New York: Basic Books.

Saffran, J.R., R.N. Aslin, and E.L. Newport. 1996. Statistical learning by 8-month-old infants. *Science* **274**:1926–1928.

Sekiyama, K., and Y. Tohkura. 1993. Inter-language differences in the influence of visual cues in speech perception. *J. Phonet.* **21**:427–444.

Sharma, A., N. Kraus, T. McGee, T. Carrell, and T. Nicol. 1993. Acoustic versus phonetic representations of speech as reflected by the mismatch negativity event-related potential. *Electroencephalogr. Clin. Neurophysiol.* **88**:64–71.

Skinner, B.F. 1957. Verbal Behavior. New York: Appleton-Century-Crofts.

Strange, W., ed. 1995. Speech Perception and Linguistic Experience: Issues in Cross-language Research. Timonium, MD: York.

Studdert-Kennedy, M., and M. Mody. 1995. Auditory temporal perception deficits in the reading-impaired: A critical review of the evidence. *Psychon. Bull. Rev.* **2**:508–514.

Sussman, J.E., and V.J. Lauckner-Morano. 1995. Further tests of the "perceptual magnet effect" in the perception of [i]: Identification and change/no-change discrimination. *J. Acoust. Soc. Am.* **97**:539–552.

Tallal, P., S.L. Miller, G. Bedi, G. Byma, X. Wang, S.S. Nagarajan, C. Schreiner, W.M. Jenkins, and M.M. Merzenich. 1996. Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science* **271**:81–84.

Vihman, M.M., and B. de Boysson-Bardies. 1994. The nature and origins of ambient language influence on infant vocal production and early words. *Phonetica* **51**:159–169.

Vygotsky, L.S. 1979. Interaction between learning and development. In: Mind in Society: The Development of Higher Psychological Processes, ed. M. Cole, V. John-Steiner, S. Scribner, and E. Souberman, pp. 79–91. Cambridge: Harvard Univ. Press.

Werker, J.F., and L. Polka. 1993. The ontogeny and developmental significance of language-specific phonetic perception. In: Developmental Neurocognition: Speech and Face Processing in the First Year of Life, ed. B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, and J. Morton, pp. 275–288. Dordrecht: Kluwer Academic.

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

Standing, left to right:
Mike Fanselow, Randolf Menzel, Allison Doupe, Jerry Rudy, Richard Morris
Sitting, left to right: