

## Chapter 6

### The Emergence of Native- Language Phonological Influences in Infants: A Perceptual Assimilation Model

Catherine T. Best

When we hear words from an unfamiliar language spoken by a native of that language, we often have difficulty perceiving the phonetic differences among contrasting consonant (or vowel) sounds that are not distinct phonemes in our own language. Of course, we experience no difficulty with phones that are very similar to our own native-phonemes. Very young infants, however, discriminate not only the segmental contrasts of their native language but many nonnative contrasts as well. That is, they are apparently unfettered by the phonological constraints of their language environment.<sup>1</sup> Moreover, young children typically come to perceive and produce with relative ease just those phones that the language of their community uses.

It is apparent, then, that the phonology of the native language comes to exert substantial influence on speech perception and production during development. As will be discussed later in the chapter, the nature of the experiential effect on perception of nonnative segments appears to be largely an adjustment of selective attention rather than a permanent revision of the initial state of sensory-neural mechanisms. The effect of language experience is neither absolute in extent nor irremediable in adulthood, and it varies in degree among specific types of nonnative contrasts and among individuals (e.g., Best, McRoberts, and Sithole 1988; Flege 1988; MacKain, Best, and Strange 1981; Tees and Werker 1984; Werker and Tees 1984b; Pisoni, Lively, and Logan, this volume).

When and how does the language environment come to influence the perception of phones that are not contrasting phonemes in the native sound system? And how might that developmental transition provide insight into the acquisition of the native phonological system? In particular, how do young listeners come to recognize the way in which their language organizes disparate phonetic details into phonemic categories to serve distinctly linguistic functions?

These are the central issues to be examined in this chapter (see also Werker, this volume). The focus will be on segmental rather than suprasegmental contrasts, particularly on consonants rather than vowels and on nonnative rather than native contrasts. Thus, I will examine developmental reorganization in perception of phonetic differences among consonant sounds that do not signal a linguistic distinction in the native language but that infants in the early months can discriminate. An underlying assumption of this chapter is that developmental change in perception of nonnative contrasts reflects concomitant changes in the way the child perceives the sound structure of the native language, whether at the segmental or the prosodic level.

Current findings suggest that infants begin life with language-universal abilities for discriminating segmental phonetic contrasts (i.e., they are not yet perceptually constrained by the phonemic contrasts of their language environment: see note 1) but that, by the second half-year of life, listening experience with the native language has begun to influence the perception of contrasts that are nondistinctive in the native phonological system (Werker 1989; Werker and Lalonde 1988; Werker and Tees 1984a; Werker et al. 1981; see Werker, this volume). Recent findings from my own laboratory (Best, McRoberts, and Sithole 1988; Fowler, Best, and McRoberts 1990) may illuminate the means by which language-particular experience exerts its effect on perception. Specifically, the influence of the native phonological system on *adult* listeners entails the perceptual assimilation of nonnative phones to the native phoneme categories with which those nonnative-phones share the greatest similarity in phonetic characteristics. However, our findings with infants suggest that the developmental change in perception of nonnative contrasts during the second half of the first year does not yet involve the mature pattern of assimilation to native-phonemes.

Based on these findings, I propose that the native-language influence on perception of nonnative-phonetic contrasts begins with the older infant's emerging recognition that native speech sounds are structured as specific, recurring constellations, or patterns, of coordination among phonetic-articulatory gestures (e.g., the pattern of temporal coordination between the glottal-adduction gesture to begin voicing and the bilabial-release gesture that characterizes English /p/). As the older infant begins to recognize these gestural constellations in ambient speech, he or she may detect similar patterns in some nonnative phones but may be unable to do so with others. This approach to understanding the influence of the language en-



Environment on phonemic development may also help elucidate certain aspects of phonological behavior in early speech productions, as suggested by acoustic-phonetic analyses from a case study on early word imitations to be presented in the final section of the chapter. It should be noted that, while this chapter focuses on consonant contrasts, language-particular gestural constellations are assumed not to be restricted to phonemic segments but to extend also to syllables and to units of meaning in speech (e.g., morphemes).

### Language Specificity in Phonology

Languages vary in their phonological systems. Of specific interest to the present discussion, they differ in their inventories of phonemic contrasts, which are defined as segment-sized constellations of phonetic properties that have become linguistically distinctive because they are used systematically to convey differences in word meanings. For example, the lexicon of many languages, including Japanese and Korean, lacks the liquid consonant contrast /l/-/r/ found in English. Likewise, the English vowel contrast /ɪ/-/ɛ/ (as in <bit>-<bet>) is absent from Spanish, Italian, and numerous other languages. In turn, English lacks many phonemic contrasts found in other languages, such as the ejective stops found in Ethiopian Tigrinya and elsewhere, the click-consonant contrasts of African Bantu languages including Zulu, and the front-rounded vowels of German, French, and Swedish.

Even in cases where languages share a phonemic contrast, the phones involved often differ between languages in their articulatory details, that is, in the exact phonetic realization of how those phonemes are produced. To illustrate, English and French both use the phonemic contrast /b/-/p/ to mark lexical distinctions (e.g., English, <bat>-<pat>; French, <bain>-<pain>). Yet, in the English phoneme /b/, glottal pulsing (voicing) may begin either simultaneously with the release of the bilabial closure, the phone [b] in International Phonetic Alphabet (IPA) transcription, or slightly after the release, short-lag voiceless-unaspirated [p]. In the English phoneme /p/, voicing instead begins either after a longer postrelease lag, which is aspirated as an aerodynamic result of the lag in glottal adduction ([p<sup>h</sup>]), or it begins after a shorter lag (unaspirated [p]) in certain positions (e.g., following /s/ or /ʃ/).

In French, however, the phoneme /b/ is realized consistently as [b] across contexts, while /p/ is phonetically realized consistently as the short-

lag voiceless-unaspirated [p]. Thus, French fails to define a phonemic distinction between the phones [p]-[p<sup>h</sup>] as English does, whereas English fails to define a phonemic contrast between [b]-[p] as French does.

By extension, languages may share a single phoneme which nonetheless differs phonetically between the languages, as with the phoneme /r/ found in both American English (phonetically realized as the liquid-approximant phone [ɹ]) and in Spanish (realized as an alveolar tap [r] or trill [rr]). Furthermore, within a language, there are often striking dialectal differences in the phones that typically realize a given phoneme (i.e., allophones), as with /r/, which is pronounced as [ɹ] in American English but is produced as a tapped [r] or trilled [rr] in Scottish English.

Even within a dialect, a phoneme may have allophonic variants. These can vary either systematically, dependent on their phonetic context (context-conditioned allophones such as the American English /k/ realized as [k<sup>h</sup>] word-initially but as [k] following /s/ or /ʃ/) or independent of context (in free variation: e.g., final /d/ may be either released or unreleased in American English). Thus, languages and dialects use but a subset of the phonetic gestures of which the human vocal tract is capable, and they differ in how they relate those articulatory details to phonemic distinctions.

These types of language-particular phonological characteristics are known to influence speech production, the degree of the experience-based effect varying with development. The phonetic details of the native-language (L1) phonology are strongly ingrained in the production patterns of mature speakers who speak like natives not only in their choice of words but also in the accent of their speech. A corollary influence of the native phonology is that adults usually maintain an L1 accent when they learn to speak a new language (L2) and typically find it quite difficult to produce L2 with fully correct phonetic details. However, normal young children rather quickly learn to speak the language of their community with native accents; unlike adults, they usually can acquire additional languages prior to 5-6 years of age with little or no trace of accent from L1 phonology (Brière 1966; Flege 1987; Flege, McCutcheon, and Smith 1987; Oyama 1976; Tahta, Wood, and Loewenthal 1981; see review, Flege 1990).

### **Language-specific Influences on Adult Speech Perception**

But what might these observations suggest regarding the listener's perception of native and nonnative speech sounds? Given the relative ease with which children learn to speak the ambient language(s) with appropriate

phonetic detail (i.e., both L1 and L2), they certainly recognize the articulatory properties of the speech around them. Furthermore, because this occurs regardless of which particular language is being learned, we can assume that, at least at some time during the developmental process of language learning, the auditory system must be capable of physiological sensory registration of the acoustic results of the phonetic gestures employed by natural languages.<sup>2</sup> But something changes developmentally with respect to the perceptual recognition of the organization of phonetic-articulatory details within phoneme categories. Whether or not sensory registration changes as a result of auditory exposure to particular sound patterns, perceptual recognition ability certainly does change as a result of the child's listening experience with a particular language. It is the latter, not the former, type of developmental change that is of concern in this chapter, which addresses the question, what is the nature of the developmental change in speech perception?

One possibility is that the perceptual change results from experience with producing the sounds of L1. But a potential problem with that possibility is the paradoxical implication that correct performance can precede competence. Alternatively, the effect of the native language on perception could be independent of its influence on production. For instance, while developmental changes in ease of producing nonnative sounds with correct phonetic detail might result from a history of differential articulatory practice, speech perception could remain unaffected by L1 acquisition. Although existing evidence suggests there may indeed be disparities between the improvements in production and in perception of nonnative contrasts by adult L2 learners (Flege 1988; Goto 1971; Sheldon and Strange 1982), there have been no direct tests of the perception-production relation in L2 acquisition during the young child's sensitive period for learning an L2 without an L1 accent (see Flege 1987).

Regardless of the relation between perception and production in acquisition of L2, perceptual research has shown that the phonological characteristics of the native language do indeed influence the perceptual tendencies of mature language users. Monolingual speakers of languages with differing phoneme inventories and/or with differing phonetic realizations of a given phoneme show language-particular differences in perceptual sensitivities to native versus nonnative contrasts. The cross-language pattern of variations in discrimination performance on the contrasts tested have been generally consistent with the phonemic inventories of the languages studied. For example, monolingual adult speakers of English and Thai show language-appropriate differences in their perceptual bound-

aries between voice onset time (VOT) categories along synthetic-stop-consonant continua (Lisker and Abramson 1970), monolingual Korean and Japanese speakers have difficulty discriminating the English /l/-/r/ contrast (Gillette 1980; Goto 1971; Miyawaki et al. 1975), and monolingual English speakers have difficulty discriminating the Czech voiced, retroflex-alveolopalatal fricative contrast /ʒa/-/ʒa/ (Trehub, 1976) as well as several Hindi and Native American contrasts that are not used in English (Tees and Werker 1984; Werker and Tees 1984b).

It might be that the difficulties adults have with discriminating the pair members of many nonnative contrasts reflect a permanent, absolute loss of sensory-neural sensitivity to the acoustic properties of those contrasts or to their linguistic properties (Eimas 1978) due to a lack of environmental exposure during a critical or sensitive period of development (see Aslin and Pisoni 1980b). However, lack of a phonemic contrast (i.e., linguistic information) need not imply lack of exposure (i.e., acoustic or phonetic data). Several factors mitigate both against an absolute lack of exposure to the nonnative phonetic properties and against absolute sensory-neural loss. The absence of a given contrast does not assure that the environment is devoid of the crucial acoustic and/or phonetic properties of the phonemes involved since those patterns may be present in allophonic variants (MacKain 1982) or in nonspeech sounds (Best, McRoberts, and Sithole 1988). Moreover, perceptual recognition of phonetic organization in unfamiliar speech sounds can remain open to change even in adults learning a new language or a new dialectal accent, although the extent of malleability may be more limited than in early childhood (Flege 1990).

Discrimination of some nonnative contrasts may be quite good even without training (Best, McRoberts, and Sithole 1988) or with only minimal training (Werker and Tees 1984b). Even for those nonnative contrasts that are initially difficult for the listener, discrimination may become better, with some apparent limits in degree and in generalization across phoneme contexts, through extensive naturalistic conversation experience with L2 (MacKain, Best, and Strange 1981; Mochizuki 1981), through more extensive laboratory training (McClasky, Pisoni, and Carrell 1983; Pisoni et al. 1982; Strange and Dittman 1984; Pisoni, Lively, and Logan, this volume), or under listening conditions that minimize memory demands and/or phonemic-level perceptual constraints (Werker and Logan 1985; see also Carney, Widin, and Viemeister 1977; Pisoni and Lazarus 1974). It is relevant here to note that there are also individual differences in the degree of difficulty listeners have with a given nonnative contrast (MacKain, Best, and Strange 1981; Mann 1986; Strange and Dittmann

1984). Finally, even when listeners have difficulty discriminating a particular nonnative contrast, they have been shown to discriminate the critical acoustic features when these are isolated from the speech context (Miyawaki et al. 1975; Werker and Tees 1984b).

These observations indicate that the effect of the native language on perception of nonnative contrasts is neither absolute nor permanent and, hence, cannot be fully accounted for by sensory-neural mechanisms. Rather than causing a sensory-neural loss of sensitivity to nonnative distinctions, the native language most likely promotes an adjustment of attention to language-particular, linguistic characteristics of speech signals, especially when the listener is focused at the level of phonemic information (Werker and Logan 1985). Although the change appears to be attentional and somewhat malleable rather than strictly physiological and permanent, the facts remain that adults have initial difficulty with many (although, as I will show, not all) nonnative contrasts, and that there are constraints on perceptual plasticity (Strange and Dittmann 1984; MacKain, Best, and Strange 1981; Tees and Werker 1984; Pisoni, Lively, and Logan, this volume). This pattern of language-particular attunement from infancy to adulthood is consistent with the observation that perceptual learning generally involves a shift of attention away from irrelevant stimulus information, as well as an increase in the ability to discover and recognize functionally relevant higher-order patterns of stimulus organization (Gibson 1966; Gibson and Gibson 1955). How do these phonological influences of the native language on speech perception arise developmentally?

#### **Development of Phonemic Influences on Perception**

In order to understand and speak the language of their environment, infants must come to perceive the phonetic, that is, articulatory and/or acoustic, properties that define the phonological organization of that language. Such language-particular perceptual attunement is essential in guiding the child's own productions, if she or he is eventually to reproduce the speech patterns of other members of the language community. The universal developmental sequence and the intrinsically motivated nature of normal language acquisition suggest that infants are well equipped to begin making this sort of perceptual adjustment.

That suggestion is also borne out by two aspects of research findings on infant speech perception, which will be discussed in more detail below. First, studies on the discrimination of both native and nonnative segmen-

tal contrasts suggest that during infants' early months, perception of segmental contrasts is largely unaffected by the language of their community, although some native-language prosodic influences may appear earlier than do segmental influences (Best 1991a; Mehler et al. 1988; this volume; cf. Best, Levitt, and McRoberts 1991). Young infants' perception of segmental contrasts may reflect general prelinguistic abilities that are not yet constrained by the properties of their specific language environment (see note 1). This does not imply that infants possess an innate ability to discriminate all phonetic contrasts from all languages but only that perceptual successes and failures during the first few months cut across both native and nonnative segmental contrasts, revealing a pattern quite different from the native language constraints seen in adults. Second, infants' phonetic perception does begin to show clear influences of the native language at least in the second half-year.

What information do infants initially perceive in speech sounds, and what are they beginning to recognize about higher-order organization in speech as they approach acquisition of their native language around the end of the first year? As they become language users, infants must move from detecting only general information in speech (e.g., simple phonetic properties) to recognizing and producing various language-particular functional elements (e.g., words and phonemes) carried in the signal.

Several key questions about this developmental transition must be addressed. Is the general information in speech that is perceptually accessible to the young infant linguistic or nonlinguistic in nature? If the infant initially perceives nonlinguistically, how does the developmental shift to perception of linguistic information take place? And finally, regardless of whether the initially detected information is linguistic or nonlinguistic, how does the infant come to recognize language-particular structural organization in the speech she or he hears? The next section will summarize the primary contemporary theoretical views and empirical investigations regarding the initial state of infant speech perception and the emergence of language-particular influences.

## **Development of Infant Speech Perception**

### **Theoretical Perspectives**

There are three general theories of speech perception, all of which carry implications about perceptual development. One view differs from the other two in its assumptions about the nature of the information that the perceiver initially apprehends in speech, that is, it assumes a different

immediate object of perception, a different sort of perceptual primitives. That approach, referred to here as the psychoacoustic theory, posits that the immediate object of speech perception, and hence of the infant's perceptual learning, is the *proximal stimulus*, or the raw acoustic components into which the speech signal is assumed to be decomposed by the auditory periphery. This view assumes that the perceptual primitives for speech perception are an array of intrinsically meaningless, simple acoustic features, such as spectral distribution patterns, bursts of bandlimited aperiodic noise, and temporally defined silent gaps, into which the speech signal can be analyzed (Aslin, Pisoni, and Juszcyk 1983; Diehl and Kluender 1989). The psychoacoustic primitives are thus analogous in nature to the simple, two-dimensional visual features, such as edges, lines, angles and spatial frequency components, that are often described for instantaneous two-dimensional retinal images of visible patterns.

By contrast, both the current motor theory (Lieberman and Mattingly 1985, 1989; Mattingly and Liberman 1988) and the ecological theory of speech (Best 1984; Fowler 1986, 1989, 1991; Fowler, Best, and McRoberts 1990; Fowler and Dekle 1991; Fowler and Smith 1986; Rosenblum 1987; Studdert-Kennedy 1985, 1986a, 1989, 1991) argue that the immediate objects of speech perception are the *distal events* that shaped the signal (see Gibson 1966, 1979). These two views assume that the perceptual primitives for speech perception are the articulatory gestures of the vocal tract, that is, the formation and release of constrictions by diverse articulators at various positions along the vocal tract (see treatments of articulatory phonological theory by Browman and Goldstein 1986, 1989, in press). Speech signals directly provide articulatory-gestural information because their complex, time-varying patterns are lawfully shaped according to the principles of acoustic physics (Fant 1960) and by the physical structure of the vocal tract and its dynamic gestures (e.g., bilabial closure, velum lowering, and glottal opening). Gestural information, then, is present in the complex organization of the speech signal as it changes over time, certainly to no less an extent than pure acoustic features are present in the signal. The motor theory and the ecological view both assume that it is the gestural information that is directly extracted from speech signals and that this information is not built up from an analysis of simple acoustic features. Thus, these views require no intervening mental step to translate raw acoustic features into gestural patterns.

The point of contention among the theories with respect to the perceptual primitives for speech perception, then, is whether the information that the perceiver extracts directly from speech is comprised of pure,

simple acoustic features or rather is comprised of dynamic articulatory patterns. The psychoacoustic approach assumes that the acoustic pressure wave is decomposed into simple, meaningless features, which serve as the immediate object of perception. However, the motor theory and the ecological view regard the acoustic waveform as but one of the energy media (along with the dynamic optical patterns of visible articulations and even the haptic patterns of manually felt articulations) that are shaped by and carry information about distal vocal tract gestures, which are the immediate objects of perception.

As background for the remainder of this chapter, a brief overview will be given of some of the primary differences and similarities of these three models. An in-depth comparative analysis of the three models, however, is beyond the scope of this chapter. A number of existing sources provide detailed treatments of each of the theoretical positions and of the debates among them; readers interested in more extended discussions of the logical and empirical grounds for each viewpoint are directed to Diehl and Kluender (1989) for an examination of the psychoacoustic approach, to Liberman and Mattingly (1985, 1989) for the presentation of the motor theory, and to Fowler (1986; see also her reply to Diehl and Kluender 1989); and Best (1984) for discussions of the ecological account, which are based on Gibson's (1979) ecological theory of perceptual systems in general.

The three speech perception theories differ with respect to how the primitives of perception are assumed to be related to the linguistic entities represented in the speech signal, for example, phonemic segments. According to the psychoacoustic perspective, the infant must ultimately learn to associate combinations of acoustic features, which are intrinsically meaningless and nonlinguistic, with the linguistic entities of words and phrases (meaningful units), as well as with the syllables and phonemes (structural units) that may be recombined to convey different meanings (see Jusczyk 1981, 1986, this volume; Pisoni, Lively, and Logan, this volume). Thus, the infant must develop auditory templates or prototypes that become paired associates of abstract linguistic entities.

The ecological and the motor-theory perspectives assume, alternatively, that the infant must discover which particular temporospatial constellations of articulatory gestures are employed as specific linguistic elements (words, phonemes, etc.) in their native language, such as the temporal relation between bilabial closure and glottal opening at the beginning of American English words like /peak/ and /pat/. Unlike the motor theory, however, and analogous to the psychoacoustic view, the ecological view



assumes that the information young infants initially detect in the speech signal is nonlinguistic. According to the ecological approach, the sort of distal articulatory information detected by the prelinguistic infant is initially devoid of linguistic relevance for them. Young infants, and indeed other animals, presumably also detect analogous distal event information in other environmental sounds and sights (see Best 1984; Fowler, Best, and McRoberts 1990; Studdert-Kennedy 1986a).

For example, recent studies show that young infants can recognize lawful temporal macrostructural (rhythmic) as well as microstructural (object composition) commonalities between the sights and sounds of single versus multiple marbles being turned back and forth inside plexiglass containers (Bahrck 1987). They come to recognize the intermodal relations on the basis of physically lawful relationships between the objects/events and the corresponding sounds, and they show no evidence of learning intermodal matching when sight and sound are only arbitrarily associated (Bahrck 1988). Also, adults' perceptions of auditory nonspeech events involving steel balls rolling down two-part runways with differing slopes are consistently determined by the dynamic properties of the distal events in ways that cannot be reconciled by psychoacoustic principles such as auditory contrast (Fowler 1991).

With respect to speech, the human child differs from other animals because, at some point in development, she or he begins to discover sound-meaning correspondences between higher-order patterns of articulatory gestures and specifically linguistic functional elements, such as referents for objects, events, and people and their interactive relationships. The child discovers that mature speakers organize their articulatory gestures into systematic, recurring constellations in order to convey different meanings. These gestural constellations are the physical instantiations of the multiple, nested levels of linguistic organization in speech (phonemes, morphemes, words, phrases, and sentences), all of which are specific to the language environment (see also Best 1984).

Thus, the ecological view and the psychoacoustic view put forward the notion that the basis for speech perception is initially nonlinguistic in nature, sharing common ground with the perception of nonspeech sounds, and that speech perception must then shift developmentally to a linguistic basis. However, the two views obviously differ with respect to the nature of the nonlinguistic information the infant is presumed to derive initially from speech and as to the means by which the infant is presumed to make the developmental shift to a linguistic basis for speech perception.

The current motor theory (Liberman and Mattingly 1985, 1989; Mattingly and Liberman 1988) differs from the ecological theory in two ways. It assumes that even the simple articulatory gestures that infants perceive in speech are linguistic in nature, specifically that they are phonetic, and that these gestures are detected from the outset via a biologically specialized language module in the human brain that is independent from the general neural mechanisms involved in the perception of all other, non-linguistic information. The detection of distal event information is a characteristic of specialized, or closed, modules (see Fodor 1983) but not of general, or open, modules, such as the general auditory system that handles perception of other types of sounds. Here, again, the motor theory differs from the basic assumptions of both the psychoacoustic model and the ecological approach. The motor theory assumes that perception of nonspeech auditory patterns presumably proceeds, more or less, in the manner described by the psychoacoustic view rather than that described by the ecological view, that is, nonspeech auditory perception is assumed to begin with detection of the proximal acoustic properties at the auditory periphery, which does not involve detection of distal event information. According to the motor theory, the task of the uniquely human phonetic module is to relate the articulatory gestures detected in speech to the more abstract phonological units and to translate abstract structures, such as words, into neuromotor commands for producing specific utterances.

The view of the development of speech perception and production that is taken in this chapter follows the ecological approach. It begins with the premise that the articulatory-gestural properties of ambient speech serve as the primitives for the infant's task of learning to use speech as a tool for communicative purposes within a particular language. Analogous to the way we tend to perceive the characteristics of a physical tool (e.g., an adze) in terms of possible goal-related actions with that tool (i.e., its affordances: Gibson 1966, 1979), learning about speech must entail a link between perception and action (speech production) in the context of affordances—outcomes that the speaker-listener perceives can be accomplished by vocal communication, such as shared games or positive emotional interactions (see also Dent 1990)—in order for the child to become a speaker of a particular language. Thus, the immediate object of the infant's perception of speech is the pattern of articulatory gestures that shaped the signal. These gestures are the first and foremost properties that the infant must recognize so that she or he can come to use the vocal tract as a tool for language-specific communicative purposes.

The notion that gestural information is the basis of infant speech perception is supported by evidence that 4- to 6-month-old infants match visual and auditory speech patterns in bimodal perception studies (Kuhl and Meltzoff 1982, 1984; MacKain et al. 1983) and that even younger infants show perceptual compensation for coarticulatory information in speech (Fowler, Best, and McRoberts 1990) according to the same pattern found in adults (Mann 1980, 1986). The latter reports considered several likely psychoacoustic explanations for the perceptual shifts observed in differing coarticulatory contexts, including auditory contrast and critical band effects but rejected those accounts on both logical and empirical grounds in favor of an articulatory basis for the perceptual patterns.<sup>3</sup>

As for the bimodal speech perception findings, the psychoacoustic account that has been offered is that the match between the optical pattern and the acoustic pattern could have been learned by association. Although no published studies have tested the association-learning account in infants, two sets of findings with adults mitigate against an associationist explanation. Specifically, synchronized, but incongruent, audiovisual speech stimuli often yield singular phonetic percepts that do not correspond to either the audio or the video signal, and could not have learned by direct association (MacDonald and McGurk 1978; McGurk and MacDonald 1976; cf. Massaro 1987).

Even more damaging to the associationist account are the recent findings of Fowler and Dekle (1991). In that study, subjects were tested for the McGurk-MacDonald type of cross-modal phonetic percepts under incongruous auditory-haptic presentations. The subjects' fingers rested against the silently moving lips of an unseen face while they were played a synchronous, but phonetically incongruent, auditory token. They were also tested bimodally with incongruous auditory-orthographic presentations. Haptically felt information about lip movements during speech is the lawful outcome of those speech movements, just as the acoustic speech signal is the lawful outcome of those same articulatory gestures, whereas orthographic symbols relate to phonetic segments by convention, that is, by arbitrary association. Although the subjects had no prior experience with haptically felt speech and, hence could not have formed previous auditory-haptic associations, they had many years of reading experience founded explicitly on repetitive arbitrary association between phonemes and specific orthographic symbols. Nonetheless, these subjects showed cross-modal phonetic percepts akin to those found in the McGurk and MacDonald studies only under the auditory-haptic condition. Their per-

cepts in the auditory-orthographic condition were determined directly by the auditory stimuli and were completely unaffected by the simultaneous orthographic presentations. Thus, the results clearly run counter to the associationist account and support the ecological account that speech perception is based on gestural information.

As discussed earlier, the distal-gestural patterns of utterances are organized at multiple linguistic levels. But, the ecological view assumes that these linguistic levels of organization can be detected only by perceivers who have become familiar with ambient speech and have begun to discover its affordances for conveying meaning, such that they can recognize the invariant and contrastive properties of its language-particular gestural constellations (see Dent 1989 for an ecological account of semantic and syntactic development). Accordingly, the emerging influence of the language environment on speech perception involves a shift from the detection of nonlinguistic information about simple gestural properties of speech to the detection of higher-order and functionally linguistic coordinations among articulatory gestures. By hypothesis, this shift begins as the child discovers, during the final quarter of the first year, that contextually defined references to real world objects/events (meanings) repeatedly co-occur with specific patterns of intergestural constellations in spoken words and phrases. Thus, ecological theory assumes that these emergent properties of speech gestures are themselves the linguistic entities rather than assuming the linguistic entities to be abstract, static mental representations. Language is composed of dynamic action patterns—whether spoken, manually gestured, or written—whose function is to afford speakers and listeners a means by which to communicate about actual or potential activities in which they may wish to engage, such as to indicate rules about a game to be played or to collaborate on a joint project.

A major appeal of the ecological approach to speech development, therefore, is its assumption that perception and production share a common metric of information—the articulatory gestures of the human vocal tract. This perception-production link is crucial to the language-learning child, who must come not only to recognize the patterns of native words across acoustically diverse productions by widely differing speakers but also to produce reasonable approximations of those patterns. By the ecological account, no translation is needed between perception and production because they are informationally compatible.

The psychoacoustic approach assumes informational incompatibility between perception and production. The auditory percepts are specified in

acoustic but not motoric terms, and the production patterns are specified in motor-control but not auditory terms, so acoustic  $\langle \rightarrow \rangle$  motor translation is required. Because learning a native language requires that perception of the acoustic signal and self-production of speech be, or become, informationally compatible, the child of the psychoacoustic approach would either have to learn (i.e., construct) algorithms for linking perception and production or else the translation routines would have to be innate. In either case, cognitive operations would have to transform acoustic and motor-control parameters to some common abstract form of information, presumably linguistic in nature, that is, phonetic categories or features. Thus, behind the superficial acoustic cues or templates or prototypes must lie abstract mental representations of linguistic entities.

Most existing versions of the psychacoustic model do not address the acoustic  $\langle \rightarrow \rangle$  motor translation issue. But the premises of the model would seem to mandate that the translation routines be learned associatively (Diehl and Kluender 1989) rather than determined innately, or otherwise, the psychoacoustic approach would paradoxically mirror a central tenet of the original motor theory (Liberman et al. 1967).

The implication of associative learning for auditory-motor translation routines brings us back full circle to the problem of informational incompatibility between auditory perception and motor production. It is by no means a trivial matter to accomplish the necessary bootstrapping from one form to the other, especially given the problem that auditory feedback from self-produced speech could not be expected to provide guidance to motor control for the very reason that it is informationally incompatible at the outset (see Fowler and Turvey 1978). If the auditory and motor information are incompatible, it is not clear how the child would decipher which properties of the auditory signal are, or ought to be, associated with which aspects of motor production, or how she or he would be able to evaluate whether the correct associationist inference had been formed. These basic logical problems with the implications of the psychoacoustic approach are among the primary reasons that I reject that approach and will give no further consideration to it in my subsequent discussion of early developmental changes in cross-language speech perception.

The current motor theory, like the ecological approach, postulates articulatory gestures as the primitives of speech perception (Liberman and Mattingly 1985, 1989; Mattingly and Liberman 1988). The motor theory requires no cognitive translation from acoustic cues to phonetic elements, which are directly and precognitively perceived as the distal gestures of the speaker's vocal tract articulators. However, unlike the ecological ap-

proach, the motor theory proposes that a specialized phonetic module is needed, in part to translate the abstract gestural patterns of words, phonological elements, and so forth, into neuromotor commands for speech production.

The ecological approach avoids the need for translation between different forms of information either in perception or in the perception-production link and, thus, has the benefit of parsimony. What the infant needs at the start is simply the general property of perceptual systems as described by Gibson (1966, 1979): that they are organized for the detection of information in stimulation about the distal objects and events that shaped the energy patterns reaching the perceiver, particularly with respect to information about the actions that those objects/events may afford the perceiver. By this general definition, human infants do not differ from other species in their general approach to perceiving speech or other events.

But, ultimately, we do differ from other species with respect to the specific affordances that speech holds for us. We alone possess the apparatus for producing the gestures of speech (e.g., Lieberman 1975). The few avian species that can approximate some of its acoustic patterns do so by quite different vocal mechanisms. More importantly, their imitations are not used meaningfully, and they are holistic, failing to reorder or recombine phonetic, syllabic, or lexical subunits to create new utterances. In fact, so far as we know, no other species systematically varies the sequence of discrete elements of meaning and/or structure in order to change the intent of their communicative messages, as the phonological and syntactic organization of human languages allow us to do. Even the apes who have been taught to use languagelike systems of manual, or visual signs, to communicate with humans (Gardner and Gardner 1973; Premack 1971) have not mastered the grammatical functions of word order or the use of the closed class, which are both crucial characteristics of syntactic systems and hence of true language (Aitchinson 1983; Terrace et al. 1979). Even at the level of phonetic perception, there is recent evidence that monkeys do not organize speech categories around prototypic exemplars as human adults and infants do (Kuhl 1991). Thus, in the context of communicative interactions, human infants move beyond other species as they discover, within the context of human social and communicative interaction, the multiple, interlocking levels of linguistic organization in speech that are carried by language-particular constellations of articulatory gestures.

Having established the theoretical background, we will turn next to a brief review of empirical findings regarding the influence of the language

environment on infant speech perception. We will attempt to discover what information young infants perceive in speech sounds and how this changes as they become attuned to the native language. Our focus is on recent laboratory findings on perception of nonnative speech contrasts, that suggest that, although the language environment has begun to influence speech perception before the end of the first year, the older infant's perception is not yet fully organized according to the phonological system of the native language that guides adults' perceptions of nonnative sounds. But, first, we must summarize earlier findings on the general characteristics of infant speech perception, and on possible developmental influences of the language environment.

### Language-particular Developments in Infant Speech Perception

Numerous studies in the past two decades have examined young infants' abilities to discriminate a wide variety of phonetic contrasts during the first half year of life, although almost none of these have examined infants under 2 months of age (see for reviews Jusczyk 1985; Kuhl 1987). The results have indicated that, in general, 2- to 6-month-old infants can discriminate between-category differences for most synthetic or naturally produced phonetic contrasts on which they have been tested. The between-category pairings that these young infants have discriminated, as well as those with which they have had difficulty, include both native and nonnative contrasts (e.g., Trehub 1973, 1976). Although it would be an overstatement to claim that young infants possess an innate ability to discriminate all phonetic contrasts from all languages, nonetheless, their speech perception abilities are broad, are relevant to many phonetic category distinctions, and are apparently general across languages (i.e., show cross-language similarities) rather than being biased by their specific language environment as those of adults are.

The few segmental contrasts that have been reported as difficult for young infants to discriminate are consistent with the notion that general, rather than language-particular, abilities underlie early speech perception because they have involved both native and nonnative contrasts. In several studies, English-learning infants under 6 months have failed to discriminate certain English fricative contrasts: /f-θ/ (Eilers, Wilson, and Moore 1977) and /s-z/ (Eilers 1977; Eilers and Minifie 1975). However, they have discriminated other fricative contrasts, such as /s-/ʃ/ (Kuhl 1980). While some published reports have found discrimination by infants 6 months and younger of such difficult fricative contrasts as /f-θ/ (Kuhl 1980; Levitt et al. 1988) and /s-z/ (Eilers, Wilson, and Moore 1977; Eilers,

Gavin, and Oller 1982), even the latter findings have suggested that those particular contrasts may be more difficult for young infants to discriminate than are other contrasts such as the stop-place distinction /ba-/da/.

As for nonnative contrasts that may be difficult for infants under 6 months to discriminate, English-learning infants have failed to discriminate some acoustic voice-onset-time (VOT) distinctions in the range of the nonnative Spanish and Thai prevoiced-voiced stop contrasts (Eimas et al. 1971; Eimas 1974). This failure was of particular interest in light of two reports that such contrasts are discriminated by young infants from language environments that employ prevoiced-voiced stop distinctions: Guatemalan Spanish 4- to 6-month-olds (Lasky, Syrdal-Lasky, and Klein 1975) and Kikuyu 2-month-olds (Streeter 1976b). Unfortunately, ambiguities in these studies preclude a straightforward conclusion as to whether speech perception abilities in this early developmental period show general versus language-particular constraints (see also MacKain 1982).

Two important general limitations of all these studies was their failure to assess directly for age-related changes in the perception of native versus nonnative contrasts or to compare infants from different language environments with identical stimuli and tasks. In order to conclude that a language-particular developmental change in perception has taken place, infants from a single language environment would have to show similar responses to native and nonnative contrasts at a young age and then show a language-relevant difference in response to the two contrasts at some later age. Alternatively, support for a language-particular influence would be obtained if infants from each of the two language environments showed equivalent levels of discrimination at some younger age on a contrast that is present in one but not the other language and then yielded language-relevant differences at some later age. In either case, a lack of language differences across ages would suggest that language-particular learning had not yet clearly affected perception.

There are also more specific problems in interpreting the findings. In the reports on the Spanish- and Kikuyu-learning infants, the subjects failed to discriminate the acoustic VOT contrast actually present in productions by adults of their language community. Instead, they discriminated some other nonnative prevoiced VOT contrast. Nor has the issue been clarified by another report that young English-learning infants may be able to discriminate some acoustic VOT distinctions in the prevoicing range (Eimas 1975). The latter study employed VOT differences that were much larger than the intervals used to test for category boundaries in the Spanish-



and Kikuyu-learning infants or in other regions of the acoustic VOT continuum.

In addition, the use of computer-synthesized acoustic VOT continua in all these studies may pose a problem with respect to the voicing categories of the adult languages. Articulatory VOT distinctions among stop categories actually result in multidimensional acoustic differences between categories, but the synthetic continua manipulated only the timing of acoustic onset of periodicity following stop release. This obviously would be problematic if some property other than acoustic VOT per se were the actual or primary source of perceptual information for native listeners. Indeed, even native adults' perceptual boundaries with synthetic VOT stimuli fail to correspond to the voicing categories found in Spanish (Lisker and Abramson 1970) or Kikuyu productions (Streeter 1976a), suggesting that acoustic VOT is not the primary perceptual information that distinguishes the prevoiced-voiced categories for them (Lisker and Abramson 1970). And the possibility that even native voiced-voiceless stop distinctions are discriminated by infants on the basis of some other acoustic property besides timing differences in acoustic VOT is suggested by the failure of English-learning infants to discriminate synthetic /du-/ /tu/ stimuli. These stimuli differed only in acoustic VOT and lacked the F1 transition cut-back cue that had been confounded with VOT in other synthetic speech studies (Eilers et al. 1981).

Thus, there is only sparse, equivocal evidence of any language-specific influences on the perception of consonant contrasts in infants under 6 months. The weight of empirical findings favors the view that, during their first half-year, infants possess only general abilities to discriminate many, though not all, consonant contrasts from both native and nonnative languages. This characterization does allow that some phonetic contrasts may be easier than others for young infants to discriminate. It assumes only that such variations are not yet constrained by the infant's specific language environment. However, the possibility remains open that improvement may occur even during these early months for perception of other properties available in the native language environment, that is, for the beginnings of some language-particular learning, such as global prosodic properties of the native language (Mehler et al. 1988).

Other evidence indicates that the native language does begin to influence perception of phonetic contrasts during the second half-year of life. The emergence of language-particular perceptual effects during this developmental period would be consistent with general observations that infants generally start to produce their first words by the end of their first

year and that they begin to understand words even earlier, by between 8 and 10 months. Both of these observations imply the development of sensitivity to the sound patterns of native words.

The first such studies with older infants suggested that 6- to 8-month-olds from different language environments may differ in their discrimination of native and nonnative-phonetic distinctions. Those studies examined Spanish- and English-learning infants' discrimination of synthetic versions of a prevoiced-voiced stop distinction found only in Spanish and a voiced-voiceless stop distinction found only in English (Eilers, Wilson, and Moore 1979; Eilers, Gavin, and Wilson 1979). These studies also tested the discrimination of naturally produced distinctions between the tapped versus trilled /r/ ([r]-[r]) found only in Spanish, the fricative-voicing contrast /s/-/z/ found only in English, and the Czech fricative-place contrast /z/-/ʒ/ found in neither Spanish nor English (Eilers, Govin, and Oller 1982). The results suggested that Spanish-learning infants discriminated the Spanish voicing and /r/ contrasts, while the English-learning infants showed marginal or no discrimination of these contrasts. However, both groups discriminated the English and Czech contrasts, with the Spanish-learning infants performing no worse than the English-learning infants on the former and actually performing significantly better on the latter. The authors concluded that the language environments of the two groups of infants differentially affected their discrimination of the cross-language contrasts.

These findings suggest a possible language-particular influence on phonetic perception at 6-8 months. Although some concerns about methodological and interpretive difficulties were raised by Aslin and Pisoni (1980b), Jusczyk, Shea, and Aslin (1984), and MacKain (1982), the authors have rebutted many of the criticisms (Eilers, Gavin, and Wilson 1980; Eilers et al. 1984). However, some ambiguities remain. As with the studies of younger infants, age change was not assessed. One report by Aslin et al. (1981) on discrimination of nonnative synthetic prevoiced/voiced stop contrasts by English-learning infants between 6-12 months does little to help resolve the issue. Subjects were not assessed for age changes in perception, and only a very small number of the subjects who began the study completed the prevoiced/voiced tests, their results showing wide variations in boundary positions tested across rather wide VOT intervals (e.g., VOT differences of up to 70 msec, as compared to the 20 to 30 msec VOT intervals used in other studies with Spanish infants).

An alternative explanation of language group differences at a single age in the Eilers's studies might be that the Spanish-learning infants are simply

better discriminators overall than the English-learning infants for some reason other than language experience itself. Indeed, the Spanish infants discriminated the English-voicing contrast as well as the English-learning infants, and they discriminated the nonnative Czech contrast significantly better than did the English-learning infants. The authors suggested several additional factors, both linguistic and nonlinguistic, that might account for the Spanish infants' high performance on non-Spanish contrasts. Spanish may provide more or better phonological analogies of those nonnative contrasts than English does, the bilingual Spanish-English environment to which the Spanish infants were exposed may have aided them in discriminating the English contrast, and/or the English voicing contrast may be acoustically salient even to infants who have not been exposed to it.

More recent findings from Janet Werker's lab and from my own lab are consistent with the idea that general rather than language-particular abilities underlie discrimination of many segmental contrasts at 6–8 months. Of greater interest, however, is the related developmental finding that unequivocal language-specific changes in perception of nonnative contrasts certainly have begun to appear around 8–10 months of age and are strong by 10–12 months. Using a version of the conditioned head-turn technique, Werker and colleagues (1981, 1984, 1988) presented English-learning infants at 6–8, 8–10, and 10–12 months with an English stop-place contrast (/b/-/d/) and with the following nonnative contrasts: Hindi dental-retroflex stops /ɖ/-/ɗ/, breathy voiced-voiceless dental stops /d<sup>h</sup>/-/t<sup>h</sup>/, Thompson Salish (Native American) ejective velar-uvular stops /k'/-/q'/. In their several studies, the authors have used both natural CV syllables and synthetic continua, as well as both cross-sectional and longitudinal developmental designs. Yet, regardless of the variation in stimuli and experimental design, the results have been remarkably consistent. At 6–8 months, the infants discriminated not only the native contrast but also all nonnative contrasts, while at 10–12 months they showed significant discrimination only for the native contrast. Results for the 8–10-month-olds showed intermediate levels of discrimination for the nonnative contrasts. For comparison, several Hindi and Salish infants tested at 10–12 months showed good discrimination of their native contrasts, the same contrasts on which the oldest English-learning infants had failed. On the basis of these findings, Werker hypothesized that the reorganization in infants' perception of nonnative contrasts by 10–12 months of age reflects the emergence of the native phonological system.

Werker's findings are exciting because they suggest that language-particular perceptual reorganization corresponds to the period during

which infants are beginning to comprehend words and to establish a receptive vocabulary. It also corresponds to the period during which many infants move from producing only reduplicated babbles, in which a single syllable is repeated several times, to incorporating variations in consonant and vowel-like elements within their multisyllabic babbles (Oller 1980; Stark 1980).

Moreover, at the same time, infants are making the transition from Piaget's third sensorimotor substage of secondary circular reactions to the fourth substage of means-ends differentiation. This cognitive shift suggests the possibility that while younger infants at the secondary circular-reaction stage may attend to and discriminate among speech sounds because of their interesting sound properties, after shifting to the means-ends stage infants may become more interested in speech sounds as functional means that can be directed toward communicative ends. Thus, concomitant with the cognitive transition, there may be a shift toward perceiving speech sounds as members of functional linguistic categories in the infant's own language community. This could be expected to have adverse effects on perception of nonnative speech sounds. Recent findings from Werker's group indicate that this cognitive transition to means-ends differentiation is indeed strongly associated with the developmental decline in perception of nonnative contrasts (Lalonde and Werker 1990). The timing of the cognitive shift for individual infants neatly predicted their loss of discrimination for the nonnative contrasts.

The Werker hypothesis about phonological influences at 10–12 months of age is intriguing, but it suggests a different developmental pattern in perception than is provided by recent accounts of phonological development based on the productions of older, language-learning children. Specifically, it implies that the infant begins constructing his or her perceptual map of the native phonological system with phonemic segments as the basic building blocks. Presumably, the rapid expansion of the child's receptive vocabulary during the second year would then consist of words built up from phonemic segments. Yet, researchers in child phonology have instead argued that the phonological system and phonemic contrasts are differentiated out of larger linguistic units, emerging only *after* the child has acquired a sizable vocabulary rather than preexisting as the building blocks for the larger units.

Recent findings from young children's speech have indicated that the earliest linguistic units are morpheme-, word-, or even phrase-sized (Ferguson 1986; Macken 1979; Macken and Ferguson 1983; McCune and Vihman 1987; Menn 1971, 1978, 1986; Waterson 1971). From these

global units, first syllables, and subsequently phonemes and phonemic contrasts, are only later differentiated (Lindblom, MacNeilage, and Studdert-Kennedy 1983), both in production (Goodell and Studdert-Kennedy 1990; Nittrouer, Studdert-Kennedy, and McGowan 1989) and in perception (Best, 1984; Studdert-Kennedy 1981, 1986b, in press), most likely in response to the pressure that vocabulary growth exerts on the organization of the mental lexicon (Studdert-Kennedy 1987, 1991). If we extend this reasoning to the emergence of language-specific influences in infant speech perception, then we would expect language-particular reorganization in infants' perception of speech to be initiated not by the recognition of phonemic contrasts but rather by the discovery of global patterns of gestural organization in native utterances, from which phonemes may later be differentiated.

In either case, the cross-language findings with infants raise important questions. Would the developmental pattern hold for all nonnative contrasts? In particular, might there be some types of nonnative contrasts that would remain discriminable because of their specific similarities to, or differences from, native contrasts? These questions can actually be traced to several underlying theoretical questions. By what means does the mature listener's phonological system affect the perception of nonnative sounds? And what can the answer to this question suggest to us about the development of the phonological system—the way in which the child moves from perceiving general information to discovering language-specific organization in the speech signal? Is the difference between the young infant and the 10- to 12-month-old best characterized as a transition from prelinguistic to phonological or by some other sort of perceptual reorganization?

These are the questions I have addressed in my recent research on infants' and adults' perception of various nonnative contrasts, which were chosen to differ in their phonetic-articulatory relationship to categories in the phonological system of the listeners' native language. This work was sparked by a consideration of how the constraints of the native phonological system might be expected to influence the perception of nonnative-phonetic contrasts.<sup>4</sup>

### **A Perceptual Assimilation Model for Nonnative Speech Contrasts**

When presented with a speech contrast that is not employed by the native language, the mature listener is confronted with discrepancies between the properties of the nonnative sounds and those of native phonemes. How

do listeners respond to these discrepancies? We can generally dismiss the possibility that adults are unable to perceive any discrepancies. For example, listeners can detect discrepancies between familiar native-accented speech and that spoken with another regional accent or even with a foreign accent (Flege 1984). These observations indicate that mature listeners hear *discrepancies* between nonnative and native phones even though they often recognize sufficient *similarities* to familiar native phonemes to comprehend native-language utterances.

What is it about the discrepancies that the listener is picking up, and how do the nonnative sounds relate perceptually to native phoneme properties? The nature of the discrepancies and similarities can be viewed according to the three theoretical approaches to speech perception summarized earlier. The discrepancies and similarities may be perceived in terms of either articulatory properties or acoustic properties. For reasons already discussed, this chapter takes the ecological perspective that it is primarily the evidence about articulatory gestures in the speech signal that informs the perceiver. Thus, my premise is that phonologically mature listeners perceive in nonnative phones information about their gestural similarities to native phonemes. A listener will fail to detect discrepancies between native and nonnative phonemes if she or he perceives the phones to be very similar in their articulatory-gestural properties to a native phoneme category. In this case, the nonnative phones will be assimilated to the native phoneme category that the listener perceives to be most similar. Conversely, a listener will perceive discrepancies between native and nonnative phones if she or he cannot detect a correspondence between the articulatory-gestural properties of the native and nonnative phones that is even moderately acceptable. In this case, no assimilation would take place.

However, assimilation is not expected to be all or none. (Liberman et al. 1967) but consistent with subsequent evidence of above-chance within-category discrimination (Carney, Widin, and Viemeister 1977; Pisoni and Lazarus 1974), listeners should retain some degree of sensitivity to gestural variations even within native categories (Best, Morrongoiello, and Robson 1981; Grieser and Kuhl 1989; Werker and Logan 1985). Contrary to some early claims for absoluteness in categorical speech perception. Therefore, even as a nonnative phone is assimilated to the native category perceived to be most similar, the listener often recognizes discrepancies between them (i.e., recognizes that the unfamiliar phone is less than nativelike).

According to the reasoning just outlined, it follows that not all nonnative contrasts should be treated alike by phonologically sophisticated listeners. Only some nonnative contrasts should prove difficult for mature listeners to discriminate, while others should be easy to discriminate even without prior exposure or training. The perceptual variations should be predictable from differences in the patterns of gestural similarities and discrepancies between various nonnative contrasts and the properties of native phoneme distinctions. Specifically, Best, McRoberts, and Sithole (1988) have listed four patterns by which the two members of a given nonnative contrast could be perceptually assimilated to native phonemes

1. The members of a nonnative contrast may be gesturally similar to two different native phonemes, thereby becoming assimilated to Two Categories (TC type). For example, the Hindi retroflex stop /ɖ/ is likely to assimilate to English [d] while Hindi breathy-voiced dental stop /ɖʱ/ may assimilate a different English phoneme category, the voiced-dental fricative [ð].
2. The nonnative phones may both be assimilated equally well, or poorly, to a single native category, in which case they may be equally similar/discrepant to native exemplars of that Single Category (SC type). For example, both the Thompson Salish ejective velar /k'/ and uvular /q'/ are likely to assimilate to English [k<sup>h</sup>], although both will be heard as strange or discrepant from the English standard.
3. Alternatively, the nonnative pair may both be assimilated to a single native category, yet one may be more similar than the other to the native phoneme, that is, the nonnative phones may show differences in Category Goodness (CG type). For example, both the Zulu voiceless-aspirated velar /k/ and ejective velar /k'/ are likely to assimilate to English [k<sup>h</sup>], but the former should be perceived as essentially identical with English standard, while the latter should be heard as quite discrepant from it.
4. Finally, the nonnative sounds may be too discrepant from the gestural properties of any native categories to be assimilated into any categories of the native phonology and should, therefore, be perceived as nonspeech sounds, that is, they are Nonassimilable (NA type). For example, the suction-produced click consonants of southern Bantu languages are unlikely to assimilate well to any English phoneme categories.

### **Predictions of the Perceptual Assimilation Model**

The perceptual assimilation model predicts that phonologically sophisticated listeners will show near-ceiling discrimination of TC contrasts, given that the phones involved should assimilate to two different and easily discriminable native phoneme categories. These listeners should also show moderate to good discrimination of CG contrasts, which assimilate to a single native category but differ in their discrepancy from the ideal native exemplar because they can differentiate good from less-good exemplars

within the native category. However, discrimination of the CG-type contrasts is not expected to reach the high levels of discrimination found for TC contrasts because, even in the native language, between-category distinctions are better differentiated perceptually than are within-category variants. Mature listeners are also expected to have moderate to good discrimination of NA contrasts but for a different reason. In this case, discrimination performance will depend on how similar the two sounds are perceived to be as nonspeech sounds. For example, the Zulu clicks cited above may be easily discriminable if they sound like a cork popping versus fingers snapping, or else they may be only moderately discriminable if they sound like two different finger snaps. Different CG and NA contrasts may vary in discriminability due to variations in the degree of similarity, respectively, in their phonetic-articulatory properties or in their auditory properties. Finally, mature listeners are expected to show poor discrimination of SC contrasts, because the two phones assimilate to a single native-phoneme category but are equally similar or discrepant from the standard exemplar of that category.

Thus, the discrimination performance pattern for adults should be, from highest performance to lowest,  $TC > (NA \langle = \rangle CG) > SC$ . This prediction assumes strong phonological influence from the native language and is precisely the pattern of performance we have obtained with adult listeners across several experiments with nonnative speech contrasts.

It should be noted that CG and SC contrasts fall at different ends of a single dimension, in that both involve assimilation of a nonnative phone pair to a single native category. In the case of CG contrasts, discrimination is aided by the listener's recognition that one nonnative phone is more discrepant from the standard native exemplar than is the other. A corollary of this principle is that various CG contrasts may differ in the degree of discrepancy between the two nonnative phones and, hence, may vary in degree of discriminability. If the discrepancy from the native prototype is large for one nonnative item but very small for the other (a strong CG difference), discrimination will be better than if there is only a small difference in discrepancy between the two nonnative items (a weak CG difference). At the extreme of this dimension of assimilability, both members of the nonnative contrast are equally discrepant from native-category exemplars, in which case we have a SC contrast with poor discriminability.

Also note that most of the earlier-studied nonnative contrasts that have proven difficult for adults and older infants to discriminate fit the definition of SC contrasts or of weak CG contrasts, which could account for the



listeners' difficulties. In a few previous reports, adults have fared better with some nonnative contrasts even with little or no training, as in English-speaking listeners' discrimination of the Hindi voiceless aspirated /t<sup>h</sup>/ versus breathy voiced /d<sup>h</sup>/ stops (Werker and Tees 1984b) and Kikuyu-speaking listeners' discrimination of the English voiced-voiceless stop distinction (Streeter 1976a). The latter cases fit the definition of a TC contrast and a strong CG contrast, respectively.

Finally, it should be noted that NA contrasts, like CG contrasts, theoretically may vary in degree of discriminability, which will in these cases be determined by variations in salience of the *auditory* differences between pair members (Burnham 1986). Auditory rather than phonetic-articulatory differences should determine discrimination of NA contrasts because phonologically sophisticated listeners are expected to perceive them as nonspeech sounds, that is, as nonlinguistic mouth sounds or perhaps as sounds produced by similar nonvocal events.

The predictions for phonologically sophisticated adult listeners are clear. The expectations for young infants under 6–8 months are likewise clear, although different from those for adults. Specifically, young infants' discrimination performance should not differ in a phonologically consistent way according to the four assimilation types but rather should be good for most native and nonnative contrasts. To the extent that these young infants may show different discrimination levels for various contrasts, these variations should not follow the pattern described for adults but should instead be related to nonlinguistic differences in the complexity or salience to the young infant of the phonetic-articulatory distinctions involved.

On logical and/or theoretical grounds, however, there are several possible outcomes for the older 10- to 12-month-old infants, who have shown clear evidence in previous studies of a dramatic change in perception of nonnative contrasts. Their pattern of performance on the four assimilation types should provide insight into the nature of that perceptual change. One logical possibility is that advances in the infant's general cognitive/memory abilities may affect their responses to stimulus familiarity/novelty, perhaps leading to a simple heuristic in which sounds, including speech sounds, that occur in the ambient environment are recognized as familiar while those that do not occur are unfamiliar and therefore pose perceptual difficulties (but see earlier discussion and also MacKain 1982 for problems with the underlying assumptions of this reasoning with respect to the language environment). On this account, at least as regards speech perception, older infants on the verge of language acquisition

should become less sensitive to and/or interested in contrasts between phones that are absent from their environment, that is, both unfamiliar. By definition, this description would fit TC, NA, and SC contrasts, as well as at least, weak CG contrasts. Older infants should therefore show poor discrimination of all nonnative contrasts except perhaps the strong CG type, which contrasts a familiar nativelike phone against an unfamiliar one. This set of predictions I will call the *general familiarity hypothesis*.

Alternatively, the perceptual shift may be specifically linguistic in nature rather than simply being an instance of a general language-independent change in response to unfamiliar stimuli. There are several potential patterns by which a linguistically based reorganization might occur. If the perceptual shift by 10–12 months is a reflection that perception has shown a stagelike shift to becoming rule governed by the phonological system of the native language, then these older infants should show the same pattern of phonologically based discrimination performance as the adults of their language community. They should show excellent discrimination of TC contrasts, good-to-moderate discrimination of CG and NA contrasts, and only poor discrimination of SC contrasts. I will refer to this view as the *strong phonological hypothesis* to reflect its prediction of the infant's stagelike emergence into a higher linguistic level of perceptual organization that is governed by the native phonological system. Note that this approach entails the infant's recognition of the linguistic function of phonemic contrasts and other phonological rules, such as allophonic distributional constraints, which are problematic assumptions (see MacKain 1982 and discussion in the section Language-particular Developments in Infant Speech Perception, this chapter).

Two other possible reorganization patterns seem more likely, each of which would indicate a different path for the infant's developing recognition of native phones and phonemic contrasts. One of these possibilities is that older infants' perception is organized according to phonemic contrasts but that their recognition of the patterns of coordination among phonetic details for individual native-phone categories is still underdifferentiated. Thus, although they would be expected to discriminate clear between-category contrasts they may show greater acceptance of deviant tokens within a given phone category than adults do, that is, they may show lower within-category discrimination of the differences between good and poor exemplars, suggesting less refined mapping of the prototype space within the category (but see Grieser and Kuhl 1989).

This view could be referred to as the *phonemic contrast hypothesis*. Once again, the hypothesis entails the problematic assumption that infants rec-

ognize the linguistic function of phonemic contrasts. In this scenario, the older infants would discriminate TC contrasts, but because of their under-differentiated recognition of the coordinated phonetic details within individual native-phone categories, they would have difficulty discriminating CG contrasts, which entail within-category distinctions between good and less-good exemplars of a single phoneme. SC contrasts would likewise become difficult to discriminate as variants of a single native phone, although NA contrasts would remain discriminable as nonphones (i.e., as nonspeech). Thus, this view differs from the familiarity hypothesis by predicting good discrimination of TC contrasts, and it differs from both that hypothesis and the strong phonological hypothesis by predicting poor discrimination of CG contrasts.

According to the final hypothesis, 10- to 12-month-olds' perception may not be organized around pairwise phonemic contrasts as functional linguistic oppositions but rather may focus on the recognition of the patterns of gestural coordination that identify members within a given native category—a *category recognition hypothesis*. Note that the categories the infant comes to recognize need not, in fact, be confined to phonetic *segments*, but may also include larger gestural units, such as syllables and words. This last hypothesis, then, may be most compatible not only with the ecological view espoused in this chapter but also with the previously mentioned arguments that the child's earliest linguistic units are words, morphemes, and sometimes phrases (Ferguson 1986; Macken 1979; Macken and Ferguson 1983; Menn 1971, 1978, 1986; Waterson 1971), and that segmental phonology—phonemes whose boundaries are defined within a system of phonological contrasts—emerges only later by differentiation from these larger units in response to the pressure that vocabulary growth exerts on the organization of the child's lexicon (Lindblom, MacNeilage, and Studdert-Kennedy 1983; Studdert-Kennedy 1987, 1990). In other words, functional phonemic contrasts between phone categories are hierarchically more complex and abstract than are the coordinated gestural patterns that define category membership of a given utterance or phone. Nonetheless, the infant's recognition that two gestural coordination patterns differ from one another would be expected to lead to good discrimination of the two phones because of a recognition that either each phone is a clear member of a different phone category (i.e., the case of native-phonemic contrasts) or that one phone is a clear member of a given category while the other is not a good member of that category (i.e., the case of a nonnative strong CG contrast).

This final hypothesis predicts that the older infants should have difficulty discriminating SC contrasts, as well as TC contrasts, for which the nonnative phones are both unrecognizable to the infant as any native gestural-coordination patterns. Thus, while older infants could be expected to recognize many nonnative sounds as speech sounds because they can detect in them some of the general articulatory properties found in native speech, they should find it difficult to detect sufficient similarity between some nonnative-gestural-coordination patterns and the patterns they have discovered in specific native phones. This should make it difficult for older infants to discriminate not only the unfamiliar gestural coordinations seen in SC contrasts but also those TC phones whose gestural patterns deviate in many ways from even the most similar native phones. On the other hand, they should discriminate at least some CG contrasts as involving a nativelike gestural constellation as opposed to an unfamiliar gestural constellation, although they may not discriminate even those contrasts as well as adults do. However, they should perceive NA phones as nonspeech sounds (nonphones) because in them they would fail to detect even any global gestural similarities to native phones. Hence, they should discriminate NA contrasts on the basis of their nonspeech properties.

Therefore, the category recognition hypothesis differs from both the strong phonological hypothesis and the phonemic contrast hypothesis by predicting poor discrimination of some TC contrasts. It further differs from the phonemic contrast hypothesis also by predicting good discrimination of some CG contrasts and from the general familiarity hypothesis by predicting good discrimination of NA contrasts and of some TC contrasts.

### **Empirical Investigations of Perceptual Assimilation**

In the first test of the model's predictions, we assessed English-speaking adults' and infants' discrimination of a NA nonnative contrast, the Zulu apical versus lateral click contrast /ɿa/-/ʙa/ (Best, McRoberts, and Sithole 1988). This nonnative pair was expected to be nonassimilable to any English phonemes because the suction-release gesture used in them is not employed in any English phones. Nor, except for variation in laryngeal maneuvers, is it reasonably similar to any English gestural maneuvers as implosive stops are gesturally similar to plosive voiced stops or ejective stops are similar to voiceless stop. Moreover, the asymmetrical lingual release of the lateral click is not found in any English phonemes. The apical and lateral clicks sometimes do appear in isolation (i.e., without

vowels) as nonspeech "mouth sounds" in our culture, the former appearing as "tsk-tsk" sounds that indicate frustration or disapproval, the latter as a "chucking" sound used to indicate approval or to urge a horse along. These nonspeech occurrences may reinforce the American listener's tendency to perceive the Zulu clicks as nonspeech sounds.

We began by testing American adults' discrimination of the eighteen minimal-pair contrasts among the nine nonnasalized Zulu clicks (apical, lateral, and palatoalveolar places of articulation crossed with prevoiced, voiced, or voiceless-aspirated manner) in click + /a/ syllables. The AXB-discrimination task employed multiple natural tokens of each category (X was a physically different token from both A and B) and thus depended upon some degree of perceptual constancy for successful completion. There was no training on the click contrasts and just a few practice trials to orient subjects to the task. Even with this minimal exposure to Zulu clicks, the American adults, as predicted, discriminated all contrasts well above chance showing between 85–95% correct discrimination for all pairings except one. Moreover, the subjects' responses on a posttest questionnaire revealed that, as predicted, they had indeed perceived the clicks as nonspeech sounds made by release of tongue suction (e.g., tongue clucking) or as other sounds resulting from abrupt pressure changes (e.g., cork popping) rather than perceiving them as phonemic segments. Their lowest performance was 80% correct discrimination with the apical-versus lateral-voiced (short-lag VOT) pair, so this contrast was chosen for further testing with adults and infants.

To eliminate the most obvious acoustic difference between these two click categories, we equated the amplitudes of the clicks, and verified that they were still acceptable category exemplars in standard identification and discrimination tests with six Zulu listeners. A second AXB-discrimination test with a new group of American adults found discrimination to be virtually as good as before the stimulus manipulations, around 78% correct.

We then tested English-learning infants in age ranges of 6–8, 8–10, and 10–12 months (the ages examined by Werker and colleagues 1981, 1984a, 1989), ten of each age, for discrimination of this click contrast and of a control English contrast (/ba/-/da/), with test order counterbalanced in each age group. For comparison, we tested another group of adults using the infant procedure and extended the test to even older infants at 12–14 months. In each test, the subject was conditioned to fixate on a colorful slide to hear repetitions of the multiple natural tokens of the habituation syllable, which terminated whenever the infant looked away from the

slide. Following a significant decline in fixation times below a criterion habituation level for two consecutive trials, the audio presentations were shifted to the test stimulus for that nonnative contrast. To assess for discrimination, mean looking times were computed for the two trials immediately preceding the stimulus shift (habituation level) and for the first two postshift trials (response recovery).

The results of the infant study upheld the prediction that discrimination would remain high across all ages. The younger infants discriminated the category change, and moreover, the older infants or adults showed no evidence of a decline in discrimination (in fact, adults showed better discrimination than the infants in this habituation task—see also Eilers, Wilson, and Moore [1979]). Thus, the findings differ from those reported by Werker et al. (1981, 1984a, 1989) in that, although the click contrast is a nonnative distinction, older infants do not lose sensitivity to it. Our results are compatible with the claim that loss of discrimination for nonnative contrasts is not absolute and across the board but rather is due to differences in perceptual assimilation. This finding is at odds with the general familiarity hypothesis about the nature of the developmental change in perception of nonnative contrasts, but it is still compatible with the predictions of each of the other three hypotheses.

That first study had employed a different experimental procedure than Werker had used, however, so the possibility remained that the difference between her results and ours could be traced to methodological factors rather than to assimilation differences between the Zulu clicks and the nonnative-phonetic contrasts she had used. Therefore, in a second study we replicated both our findings with the Zulu click contrast and Werker's findings with her Salish ejectives, an SC contrast (Best and McRoberts 1989). We used our visual fixation procedure with new groups of 6- to 8- vs. 10- to 12-month-olds, twelve per age. Each infant was tested on both of the nonnative contrasts as well as on the English control contrast.

We again found that both age groups discriminated the NA-Zulu contrast and the English control, whereas only the younger infants discriminated the SC-Salish ejectives. Thus, the developmental difference between the two nonnative contrasts could be attributed to differences in perceptual reorganization for those types of contrast and not simply to methodological factors. This finding is again at odds with the general familiarity hypothesis but still fails to differentiate among the three linguistic hypotheses.

The next step, then, was to compare discrimination of other nonnative assimilation types. For this purpose, we examined three additional con-

trasts from Zulu. The TC contrast was a lateral-fricative voicing distinction /tɛ/-/tɛ̃/, produced with the tongue in essentially the position used for English /l/ but with a greater degree of constriction along the sides of the tongue. For adult American listeners, the voiceless lateral fricative would be expected to assimilate to the English voiceless-coronal fricatives /s/, /ʃ/, or /θ/, perhaps heard with a (devoiced) subsequent /l/ due to its /l/-like positioning of the tongue tip/blade co-occurring with fricative manner and voicelessness (see Ladefoged 1981).<sup>5</sup> The voiced-lateral fricative would be expected to assimilate to the English voiced fricatives /z/, /ʒ/, or /ð/, and/or the approximant /l/.

The CG contrast was a velar-voiceless versus ejective-stop distinction /ka/-/k'a/. The Zulu /k/ is virtually identical to English /k/ (both [k<sup>h</sup>]). But the ejective /k'/ involves a nonnative glottal gesture (rapid upward movement of the glottis during complete glottal closure) that should lead to its assimilation as a clearly less-than-ideal exemplar of English /k/. Thus, this stimulus pair constituted a strong CG contrast.

The remaining contrast was a plosive versus implosive bilabial-stop distinction /bu/-/ɓu/, originally chosen to represent an SC contrast. However, further consideration of the articulatory properties of these phones suggested that it was actually a weak CG contrast. Zulu /b/ is essentially like English /b/, but the downward movement of adducted and vibrating vocal folds for the implosive /ɓ/ differs primarily in degree from the laryngeal movement involved in English /b/. According to the strong phonological hypothesis for mature listeners, the English-speaking adults would be expected to discriminate the TC contrast nearly perfectly, the strong CG contrast somewhat less well, and the weak CG contrast most poorly but still above chance.

In the first part of the study, twenty-five monolingual English-speaking adults completed separate AXB category-identity discrimination tests on the three contrasts (as in Best, McRoberts, and Sithole 1988), each test again composed from multiple natural tokens. As predicted, adults discriminated the TC contrast with near-ceiling performance levels (~96% correct). They did slightly, but significantly, less well with the Zulu CG contrast (~88%), and discrimination performance was substantially and significantly lower on the SC contrast, although it was nonetheless above chance level (~66%).

The subjects' posttest questionnaire descriptions of the Zulu sounds indicated that nearly all subjects assimilated the TC contrast to two different English phonemes or phoneme clusters, although there was variability as to which native phonemes (or clusters) were named (s, sh, or thl versus

z, zh, zhl, or l), compatible with the gestural similarities and discrepancies from various English phonemes. Most heard the strong CG contrast as a normal /k/ versus a clearly deviant /k/ (e.g., choked or coughed). And, as expected, the majority heard the weak CG items as exemplars of English /b/. Only some of these subjects could identify a difference, in which they generally characterized one /b/ as murmured or swallowed.

Thus, consistent with the perceptual-assimilation model, adults appear to assimilate nonnative contrasts to the closest native phoneme categories, apparently on the basis of articulatory similarities and discrepancies. Furthermore, the discrimination performance pattern precisely mirrors the predictions of the strong phonological hypothesis, as we expected for phonologically sophisticated listeners. We recently verified near-ceiling performance with another TC contrast, the Ethiopian ejective-stop distinction /p'ε/-/t'ε/, which this time was assimilated virtually unanimously to English /p/-/t/, as expected from the straightforward gestural correspondence between the nonnative and native sounds (Best 1990).

In part two of the second Zulu study, English-learning 6- to 8- and 10- to 12-month-olds, fourteen per age, completed visual-fixation habituation tests for each of the three contrasts (Best et al. 1990). The predictions from all four developmental reorganization hypotheses were that 6- to 8-month-old infants would discriminate all three contrasts. The three linguistic hypotheses offered different predictions for the 10- to 12-month-olds. Analyses of the data for the 6- to 8-month-olds indicated significant discrimination across all three contrasts, consistent with predictions and with earlier findings. However, discrimination of the TC contrast by itself was marginal, while discrimination was significant for each of the CG contrasts.

However, contrary to both the younger infants and the adults, the 10- to 12-month-olds showed only marginal discrimination ( $p < .06$ ) across these three Zulu contrasts, although this age had clearly discriminated the NA-Zulu clicks in our two previous studies. Moreover, the contrast on which they showed the poorest performance—actually a small *decline* in fixation times at the stimulus shift—was the TC contrast that had proven easiest for the adults to discriminate. This age showed marginal discrimination of the strong CG contrast and nonsignificant discrimination of the weak CG contrast. Although the largest recovery in mean fixation time for either age was associated with the 10- to 12-month-olds' response to strong CG contrast, the statistical effect was hampered by high intersubject variability. This pattern of a high mean recovery paired with high intersubject variability suggests that some older infants may have detected



the category change while others utterly failed to. Planned comparisons of the 10- to 12-month-olds' discrimination pattern failed to support the strong phonological hypothesis (TC > strong CG > weak CG discrimination) or the phonemic contrast hypothesis (TC discrimination only), but they did offer marginal support ( $p < .08$ ) for a pattern compatible with either the category-recognition hypothesis or the general familiarity hypothesis (strong CG discrimination only).

Given that the two previous studies were consistent with the three linguistic hypotheses but not with the familiarity hypothesis, the full set of findings is most compatible with the category recognition account of perceptual reorganization at this age. But the category recognition hypothesis states that at least some TC contrasts should be discriminable to this age group. Why did this TC contrast pose such difficulty for the older babies? Two observations suggest a possible answer, although further research is needed to confirm it.

First, recall that the gestural properties of the TC phones (lateral voiced and voiceless fricatives) do not provide a close match to any singular English phonemes. These Zulu phones involve a lingual gesture similar, but not identical, to that found in our lateral approximant /l/, yet they also involve lingual constrictions generally similar, but not identical, to a variety of English fricatives (/s/, /ʃ/, /θ/, and their voiced counterparts). Thus, it would be understandable if the 10- to 12-month-old who is only beginning to recognize the gestural coordination patterns found in English phones has difficulty recognizing any clear similarities between the Zulu phones and particular English categories. Indeed, even the adults were quite variable in the exact patterns by which they assimilated the phones in this TC contrast to specific English phonemes.

Second, the older, and even the younger, infants' difficulty with the lateral fricatives does not appear to be a general problem with nonnative TC contrasts. We recently obtained evidence that, like adults, both 6- to 8- and 10- to 12-month-olds *can* discriminate the Ethiopian ejective TC contrast, which bears a straightforward gestural relation to a single English distinction, /p/-/t/. Moreover, both ages also discriminated the English fricative-voicing distinction /s/-/z/, which is similar to the Zulu TC-fricative distinction that the older infants failed to discriminate (Best 1991b). It is interesting to note, however, that the older infants showed marginally lower discrimination than the younger infants on the English /s/-/z/ contrast, which involves only a relative shift in voicing onset during an ongoing frication. Thus, it appears likely that the reason the older infants had difficulty with the Zulu fricative-TC contrast, but not with

the Ethiopian ejective contrast, was that the gestural properties of the ejectives relate more straightforwardly to English stops than the Zulu fricatives relate to any clear English categories for them.

Thus, the pattern of findings across studies for the older infants are at odds not only with the general familiarity hypothesis but also with both the strong phonological hypothesis and the phonemic contrast hypothesis. The performance of the 10- to 12-month-olds on NA, SC, CG, and TC contrasts across studies is most compatible with the category recognition hypothesis.

These results carry strong implications about the nature of developmental change in infants' perception of nonnative speech sounds and, by extension, offer insights about the development of the native phonological system. They suggest that, by at least 10 to 12 months of age, infants have begun to discover the gestural-coordination patterns that identify categories roughly corresponding to phones in their native language. The findings also indicate that at least some of their categories still may not be as well differentiated as those of adults and may not be as strongly organized according to the pairwise linguistic contrasts of the native phonological system.

More specifically, adults appear to assimilate nonnative phonemes to categories within the native language's system of phonemic contrasts, yielding near-ceiling discrimination performance on both of the tested nonnative TC assimilation contrasts. The older infants, however, and even the younger infants had difficulty discriminating the TC-fricative contrast, for which adults showed good discrimination but variable patterns of assimilation, while neither of the infant-age groups had any difficulty discriminating another TC-ejective stop contrast, for which adults were in perfect agreement about assimilation. Moreover, a study recently completed in my lab confirms that the Zulu TC-fricative voicing contrast continues to pose difficulty even at 4 years of age, by which time the strong CG contrast (Zulu /k/-/k'/) is discriminated consistently (Insabella 1990; Insabella and Best 1990).

The pattern of findings with the infants and its extension to 4 year olds suggest that perception of nonnative speech contrasts in relation to the system of phonemic contrasts in the native language is a relatively late achievement that probably rests on a solid foundation of knowledge about the coordinated phonetic patterns of individual native phones. The 10- to 12-month-old infants' marginal discrimination of the strong CG contrast, compared with the greater difficulty both ages showed on the TC lateral-fricative contrast, runs counter to the predictions of the phonemic-

contrast hypothesis. The initial stage of language-specific influence on perception of phonetic segments appears to involve the emerging recognition of the coordination of phonetic-gestural details *within* individual phone categories, rather than recognition of the phonetic distinctions that specify more abstract and linguistically functioning phonemic contrasts *between* categories. In keeping with the ecological theoretical perspective outlined earlier in the chapter, I suggest that the basis for infants' recognition of the language-specific properties of native and nonnative phones is the detection of evidence about the constellation of coordinated articulatory gestures that are associated with specific phones in the native language.

Drawing from these findings, I suggest the following developmental progression in language-specific perceptual learning about speech:

1. Young infants initially perceive simple nonlinguistic (articulatory and/or acoustic) distinctions in speech-sound contrasts, and this ability is not yet influenced by their language environment.
2. By at least 10–12 months, infants have begun to discover certain gestural coordination patterns of phones used in their native language. But their recognition of these patterns is still broad and underspecified, at least for some phone categories, and does not reflect the linguistic function of phonemic contrasts. At this point, they appear to detect gestural properties in some nonnative phones that are similar to the coordinated patterns they have begun to detect in native speech, but they are less able than adults to recognize the full pattern of similarities and discrepancies.
3. During the preschool years, the gestural coordination patterns of native phone categories become better differentiated, especially with respect to good versus less-good exemplars, but even by 4 years, perception may not yet be fully organized at the level of phonemic contrast per se.
4. By adulthood, and probably much earlier, perception of speech sounds involves the recognition of linguistic structure at the level of phonemic contrasts, and unfamiliar sounds are assimilated to native-phoneme categories on the basis of their articulatory-gestural similarities and discrepancies.

I have argued here that language-particular perceptual learning about speech involves the discovery of gestural coordination patterns that recur in the ambient language. It is this discovery that forms the basis of phonological development. This focus on the articulatory-gestural properties

perceived in native and nonnative speech sounds should also extrapolate to the development of phonological organization in the child's speech productions, that is, the ecological model of speech development posits that a common articulatory metric links perception and production. It follows that the child's emerging recognition of common gestural patterns in the ambient language should guide the development of language-specific phonological structure in his or her productions. In the final section of the chapter, we will examine this possibility in a case study of apparent phonological organization in a toddler's imitations of a diverse set of surface phonetic forms that realize a single phonemic contrast in American English.

### Phonological Behavior in Early Speech Production

The study reported here, based on phonetic and acoustic analyses of a toddler's imitations of a set of phonologically opaque adult targets, found evidence of apparent phonological sophistication in production of intervocalic alveolar stops at 20–22 months of age. The American English adult targets were disyllables containing medial /d/ or /t/, followed by <-er>, <-le>, or <-en>. In these contexts, intervocalic alveolar stops typically appear in normal conversation as the restricted phonetic variants flap [ɾ], nasal release [d<sup>n</sup>], or glottal stop [ʔ], (e.g., respectively, <wider> or <whiter>; <widen>; <whiten>). Although American English-speaking adults do not normally produce fully released alveolar stops in these phonetic environments, the child consistently substituted full alveolar stops for the phonetic variants actually found in the adult targets. Nonetheless, the phonetic and acoustic characteristics of the child's responses differed among the diverse phonetic target forms and distinguished between underlying /t/ and /d/.

The analyses of the target utterances and of the child's imitations suggest that sensitivity to the articulatory properties of the target utterances and/or articulatory constraints on her productions of the target phonemes in different phonetic environments provided the basis for her behavior. In particular, the acoustically diverse allophones present in the adult targets nonetheless all involved alveolar contact coordinated with a release of obstruent constriction at some point in the vocal tract.

I argue that the child's failure to imitate exactly the target utterances and the systematicity of her deviations reflect an important aspect of emerging phonological organization in her speech behavior. Specifically, the constraints provided by the articulatory information in the adult tar-

gets and those provided by the child's articulatory limitations or preferences determine how a phoneme would be realized in particular phonetic contexts and, consequently, how phonetically disparate forms may become related to a common underlying phonological category. This general line of reasoning is based on the model of articulatory phonology put forth by Browman and Goldstein (1986, 1989, 1992) according to which phonological phenomena such as epenthesis, assimilation, and reduction can be understood simply as lawful consequences of the gestural organization of utterances.

This study was prompted by informal observations of my daughter Aurora at 20 months of age when I noticed that she typically produced fully released intervocalic alveolar stops while imitating adult words in which medial ⟨t⟩ was pronounced as a glottal stop [ʔ]. Because of the complex pattern of contextual effects that produce diverse phonetic realizations of phonemic ⟨t⟩ and ⟨d⟩ in intervocalic position in American English, I reasoned that Aurora's imitative responses to the differing phonetic forms of medial alveolar stops should reveal the characteristics of incipient phonological organization for those phonemes in her speech productions (these data are a portion of those presented in Best, Goodell, and Wilkenfeld, in preparation).

Child phonology studies have typically excluded imitative responses from analysis on the assumption that imitations would closely match the phonetic details of the adult target and thus would not reveal much about the intrinsic organization of the child's phonology (Leonard, Fey, and Newhoff 1981; Leonard et al. 1978). Yet, Aurora's imitations were not exact phonetic replicas of the adult utterances. In fact, they differed *systematically* from the targets. Examining imitative responses offers several advantages over examining only spontaneous utterances: (1) demands on memory and lexical access are minimized, (2) unfamiliar and nonsense words can be presented to control for the influence of phonological idioms in familiar words (Moskowitz 1980), (3) there is no doubt about the child's intended target, and (4) the properties of both the child's production and the immediate adult target can be directly compared. In addition, the standard approach in child phonology research has included only broad phonetic transcription of the child's utterances and has not involved corollary acoustic analyses of the productions that might provide some converging evidence about their phonetic-articulatory characteristics.

For these reasons, I recorded Aurora in three sessions between 20 and 22 months as she imitated familiar, unfamiliar, and nonsense target words containing intervocalic alveolar stops that were realized as phonetically

diverse allophones. The targets were stress-initial disyllables containing intervocalic <t> or <d> produced as [t], [ʔ] or [d<sup>n</sup>]. Some words were familiar to the child and produced with familiar (American English, or AE) or unfamiliar pronunciations (Cockney English, or CE); others were unfamiliar words or nonsense words produced with AE or CE pronunciation. The medial stop was realized either as [ʔ] (e.g., AE <beaten> or CE pronunciation of <spittle>), as [t] (e.g., AE <batter>), or as the nasal release [d<sup>n</sup>] (e.g., AE <widen>). The responses were elicited in the context of a vocal imitation game at home, during which Aurora was quite willing to provide citation-form repetitions (sometimes multiple tokens) of target words that had been presented in the sentence frame "Can you say \_\_\_\_\_?" We analyzed the phonetic and acoustic properties of both my targets and Aurora's responses for direct comparisons.

Both speakers' utterances were computer-digitized and the disyllables of interest were extracted into separate files. After discarding a small number of utterances that were not acoustically analyzable due to background noise ( $n_{\text{child}} = 10$ ;  $n_{\text{adult}} = 4$ ), broad phonetic transcriptions were made for the medial consonants in both the adult's ( $n = 51$ ) and the child's ( $n = 52$ ) remaining utterances. The transcriptions were conducted blind as to the context of the preceding and following utterances. These transcriptions were independently verified by a second listener. The majority of the utterances yielded identical transcriptions from the two transcribers for both speakers (child = 70%; adult = 100%). The greater difficulty in transcribing Aurora's utterances is typical of the generally decreased reliability for phonetic transcriptions of young children that has been noted in other phonology studies. Discrepancies in the transcriptions for Aurora were resolved either by mutual agreement between the transcribers in a joint listening session (11% of her total utterances) or via tie breaking by a third expert listener (19% of her utterances).

Several acoustic measurements were also taken from the intervocalic portion of each disyllable for both speakers. Prerelease silence (from end of first vowel to release burst) and voice-onset-time (measured as the time from release-burst onset to beginning of glottal pulsing) were measured from the waveforms for tokens containing a stop-release burst. Also, total duration of intervocalic silence was measured for tokens without a release burst. The latter measure provided a fairly objective index of the duration and timing of glottal devoicing. To index the timing of supralaryngeal and/or glottal closure gestures, we measured the total duration of closure, whether voiced or voiceless, for the intervocalic consonant of all utterances. Judgments of closure onset were based on a substantial, fairly

abrupt decrease in signal amplitude and a qualitative change in voicing at the end of the first vowel (according to changes in the waveform of the pitch pulses and/or to perceptual evidence of voice quality change). Judgments of closure offset were based on the onset of periodic voicing in the second vowel of rapid increase in amplitude, and sometimes, of the presence of a release burst. To assess for any vowel-length differences associated with voicing differences in the medial stop, duration measurements of the first syllable were also taken.

According to phonetic transcriptions, the vast majority of Aurora's responses deviated from the phonetic properties of the adult targets. Her responses were predominantly fully released alveolar stops, corresponding to the phonological categories that underlie the diverse surface phonetic forms of the targets. Among Aurora's medial alveolar stops, the most frequent response to all three targets was /d/, which may reflect the relative ease of voiced alveolar stop production by children of this age in addition to, or instead of, reflecting the phonological status of the adult targets (Kewley-Port and Preston 1974; Locke 1983). Nonetheless, the proportion of /d/ versus /t/ responses and the pattern of less-frequent responses differed according to the phonological and phonetic (articulatory) properties of the targets. The voiced alveolar flap target [ɾ], which is the surface realization of medial /d/ or /t/ followed by syllabic [l] or [r] in adult AE, yielded the largest proportion of /d/ responses (~80%) and no voiceless [t]'s (the exceptions to /d/ were the voiced apical phones /t/ and /δ/). In contrast, the glottal stop target [ʔ], which is the surface realization of /t/ followed by syllabic [ŋ] in AE or CE and by syllabic [l] or [r] in CE, elicited more /t/ responses (~35%) than any other target, which approached the proportion of /d/ responses for this target (~45%). The nasal (velar) release in target [d<sup>n</sup>], which is the surface realization of an underlying /d/ before syllabic [ŋ], elicited an intermediate proportion of /d/'s (~60%). The exceptions in the latter case included /t/, the velar stops /g/ and /k/, and the glottal /h/.

The acoustic measures also indicated that Aurora both deviated from the acoustic properties of the adult target utterances and, at the same time, was systematically responsive to phonetic-articulatory differences among the surface forms of the targets. Note that Aurora's alveolar stop responses almost always included stop-release bursts, in contrast to the absence of bursts in nearly all of the adult targets—only two of the latter contained release bursts. More than half of the adult [ɾ] and [d<sup>n</sup>] targets contained no intervocalic silence, and in those that did, the silent period was only 20–67 msec in duration. In contrast, the [ʔ] targets had consis-

tent intervocalic silent periods of 50–150 msec. Consistent with the [r]-[ʔ] asymmetry in the target pattern, the intervocalic silence in Aurora's /t/ and /d/ responses to [ʔ] targets were systematically longer than those to [r] targets. This pattern is potentially consistent with either a gestural basis or an acoustic basis for Aurora's imitative responses. However, the intervocalic silent periods in her responses to the nasal release [d<sup>n</sup>] targets were even longer than those to the glottal stops [ʔ], in direct opposition to the bias shown in the adult targets. Thus, the child's responses in these cases cannot be a direct consequence of mimicking the acoustic properties of the targets. Instead, Aurora's responses to the [d<sup>n</sup>] targets more likely reflect increased difficulty in her ability to execute the heterorganic alveolar and velar gestures of a medial [d] followed by a syllabic [ŋ].

Separate examination of Aurora's prerelease silent periods versus her postrelease VOT measures offer some insight into the cause of the variations in her responses to the different targets. She produced the longest prerelease silent periods for the nasal targets (~100 msec), followed by glottal stops (~75 msec); her prerelease silences for flaps were substantially shorter (~20 msec). Thus, she appeared to use prerelease gap differences to differentiate systematically the three targets. However, she did not vary her postrelease VOT systematically among the three targets. Her mean VOTs ranged from 40–50 msec for all three targets and did not differ between the responses that were transcribed as /d/ versus those transcribed as /t/ (see also Kewley-Port and Preston 1974). Overall, these results indicate that the phonetic properties of the child's responses were achieved by means of different patterns of glottal timing than were the adult targets. The measurements suggest that she had greater control over the duration of the prerelease silent period than over the (postrelease) VOT to instantiate intervocalic stop-voicing distinctions.

We also examined evidence about the timing of closure itself, which is not adequately represented in the measures of silent intervals, especially for voiced stops. Given that children's utterances are longer than those of adults, the closure data were normalized by dividing the absolute closure duration by the length of the total utterance for both adult and child tokens. In all cases, the child's absolute closure durations were longer than those of the adult. However, the child appeared to have produced systematic differences in ratios of closure/utterance-duration for the three target allophonic categories. These values were largest for responses to targets with medial glottal stops (~.23), followed by those for the responses to nasal released targets (~.18), and smallest for responses to flap



targets ( $\sim .12$ ). The adult targets varied to a much lesser extent for absolute closure durations, and did not vary substantially in ratio of closure/utterance duration. In the adult targets, the ratios for utterances containing nasal releases ( $\sim .12$ ) were nearly identical to those containing flaps and glottal stops ( $\sim .11$ ).

In American English and other languages, vowel length differences preceding an intervocalic or final stop often distinguish between voiced and voiceless versions of the stop. Therefore, differences in the length of the vowels preceding [p] and [d<sup>n</sup>] targets may have provided the child with additional information about their voicing difference. Moreover, these vowel-length differences may also have been reflected in the child's productions. For these reasons, we measured vowel length in the first syllable of the adult and child utterances for all [p]-[d<sup>n</sup>]-minimal pairs such as *batten* [bæʔn] and *badden* [bæ:d<sup>n</sup>]. The use of minimal pairs avoided confounding intrinsic vowel-length differences with voicing-related vowel lengthening. Again, because children's utterances are longer than those of adults, the data were normalized by dividing the duration of the first vowel by the length of the total utterance. The adult targets showed vowel lengthening overall before /d/, but this pattern held only for the lax vowels /ʌ/, /ɪ/, and /æ/. Instead, the diphthongs /iɪ/ and /aɪ/ actually showed very slight shortening before /d/ relative to /t/. Aurora likewise showed overall vowel lengthening before /d/, but this held only for the diphthongs and not for the lax vowels, exactly the opposite of the pattern found in the adult targets. Thus, again, the child appears to have achieved the voicing contrast in a different manner than was provided by the adult targets rather than by simply mimicking the acoustic properties of the targets.

Overall, the results suggest some level of phonologically relevant organization in Aurora's behavior, supported by the fact that systematic production patterns appeared across familiar and unfamiliar words and nonwords (some presented with unfamiliar CE pronunciation) that involved variations in phonetic contexts and in surface phonetic realizations of the underlying phonological categories. Because the study compared the phonetic transcriptions of the adult and child utterances with several acoustic measures that were intended to provide evidence about the articulatory gestures involved, the data may offer novel insights into the way in which young children begin to organize the articulatory/phonetic details of their productions in relation to abstract linguistic categories in the native phonological system. Specifically, the findings indicate the following:

1. By the age of 20–22 months, Aurora had developed a systematic pattern of behavior that related the diverse surface phonetic forms of the medial consonants in the target utterances to underlying alveolar stops.
2. Aurora's behavior pattern was more complex than a simple one-to-one mapping from a single allophone to a single phonological category, that is, she did not simply imitate the simple acoustic properties of the targets. Instead, her behavior showed many-to-one mappings (she associated multiple surface forms to singular underlying phonemes) that nonetheless retained some articulatory/phonetic differentiation among the diverse, context-specific surface realizations of the categories.
3. The relationship between the allophones in the adult targets and Aurora's substitutions may be best understood in terms of articulatory characteristics of the targets and/or articulatory limitations on the child's productions, given the notable discrepancies between the acoustic properties of the targets and those of the child's imitations.

The complexity of these patterns in Aurora's productions is demonstrated not only by her nearly consistent substitution of alveolar stops for glottal stops, nasal releases and flaps but also by the differences in intervocalic silence and in closure intervals that she maintained among the stops she substituted for [r], for [ʔ] and for [d<sup>n</sup>].

The characteristics of the child's substitutions indicate a behavior pattern that could be considered to reflect the beginnings of phonemic organization, but an organization that is still immature and quite different from that seen in adult speech behavior. The results suggest that this behavior pattern is based on articulatory properties of the phones and phonological categories investigated. Consider that all of the intervocalic stop targets involve forward movement of the tongue tip into alveolar contact or approximation, either because of an intrinsic alveolar gesture in the target segment itself ([r], [d<sup>n</sup>], [ʔ]) and/or because of alveolar gestures in the following syllable, ([n], [l], or [r]). Accordingly, the vast majority even of Aurora's substitutions were consonants involving apical contact (including /n/ in response to [ʔ] and the linguodental /ð/ in response to [r]). She could have done otherwise—she was producing a broad array of stops, nasals, fricatives, and approximants at various places of articulation at this age, including glottal stops (e.g., <uh-oh> as [ʌʔo<sup>h</sup>]). In addition, other articulatory properties of the targets could be related to gestures other than alveolar stop maneuvers. Two of the targets ([ʔ] and [d<sup>n</sup>]) incorporate posterior vocal tract gestures (glottal stop and velar constrict-

tion) in addition to alveolar articulations, which could appropriately have been substituted by posterior consonants. In fact, of the "other" responses, those to [ʔ] and [d<sup>n</sup>] were all posterior gestures (/g/, /k/, and /h/), while the only "other" response to [r], which did not involve a posterior gestural component, was the linguodental (/ð/). Nonetheless, it is important to remember that the vast majority of Aurora's responses were indeed fully released alveolar stops.

It thus appears that Aurora was sensitive to the gestural properties of the target words, even if she did not mimic them precisely. Specifically, she appears to have been sensitive to the main place(s) of constriction in the vocal tract and to characteristics of the associated glottal gestures. Furthermore, she was able to incorporate information about those properties into her productions so that she could both relate her productions to linguistic categories in her native language and arrive at a phonetic realization for them within her articulatory limitations. Her primary difficulty was apparently a failure to incorporate into her imitations the precise temporal coordination among the supralaryngeal and the glottal gestures—the temporal phasing among discrete gestures of different articulators—that was provided in the adult targets. These findings are compatible with the ecological approach to speech development discussed earlier in the chapter and with the principles of Browman and Goldstein's model of articulatory phonology (1986, 1989, 1992).

These imitation data carry important implications about the development of phonological organization in children's speech productions, which appear to complement our findings on perceptual assimilation of nonnative phones. The pattern of systematic, context-specific articulatory variations in Aurora's productions, along with their relation to the voicing categories of alveolar stops, suggests that she related diverse allophonic realizations to common phonological categories. Her phonological categories, however, were organized differently than in adult speech production and appear to be underdifferentiated with respect to the variations in gestural coordination that are found among allophones in the native phonology. This pattern would seem most consistent with the category recognition hypothesis discussed earlier regarding developmental reorganization in infants' perception of nonnative contrasts.

## Conclusion

This chapter has examined the way in which the infant may come to relate the phonetic details of speech in his or her language environment to the

more abstract linguistic categories of the native phonological system. To this end, I have presented a model of how the infant may move from perceiving general information about phonetic contrasts in speech during the first six months of life to discovering language-particular patterns that ultimately correspond to the phonemes and phonemic contrasts that guide speech perception and production in the mature language user. This discovery, in turn, influences the properties perceived in nonnative speech sounds. I have also presented complementary information about the possible relations between the phonetic properties of adult utterances in the native language and the emergence of phonologically relevant organization in a young language-learning child's speech productions.

The line of reasoning developed in this chapter is compatible with the premise that the recognition of phonemes as specifically linguistic elements, which convey meaningful contrasts and are functionally organized within a phonological system, develops only gradually as the child builds a lexicon (see Studdert-Kennedy 1987, 1989; Flege 1990; cf. Jusczyk 1986, this volume). According to that view, phonemic segments are differentiated from words rather than being preexisting elements that are concatenated into words, and the child's phonological system emerges in accord with the principles of self-organizing systems (Studdert-Kennedy 1987, in press).

I have suggested that the means by which this phonemic development takes place in perception and production is through the young speaker-hearer's detection of information in speech about the articulatory events that produced the signal. This ecological view of speech development, which was based upon James Gibson's general ecological theory about perceptual systems, thus posits a common articulatory link between perception and production of speech. That commonality should provide obvious benefits for the young child's acquisition of a native language. The data summarized here support the proposed model of phonological development and are compatible with the ecological perspective described. I would argue that this ecological perspective on speech offers important and unique insights about the relation between perception and production in the development of spoken language. Future research must resolve the full time course of development of mature phonological organization in speech perception and production.

### **Acknowledgments**

The preparation of this chapter, as well as the original research reported here, were supported by NIH grants HD01994 to Haskins Laboratories

and DC00403 to the author. The ideas presented in this chapter have gained much clarity and detail as a result of discussions with numerous colleagues and students. I would like to extend particular thanks to Carol Fowler, Janet Werker, Doug Whalen, Winifred Strange, and Michael Studdert-Kennedy for stimulating discussions and constructive criticisms of the model as I have worked on it over the past few years. I also appreciate comments on earlier versions of this chapter by Alice Faber, Alvin Liberman, Michael Studdert-Kennedy, Jim Flege, Rebecca Eilers, and the book editors. I am especially indebted to my collaborators and research assistants: Gerald McRoberts, Elizabeth Goodell, Nomathemba Sithole, Deborah Wilkenfeld, Jane Womer, Stephen Luke, Rosemarie LaFleur, Shama Chaiken, Laura Klatt, Laura Hampel, Laura Miller, Leslie Turner, Pia Marinangelli, Ritaelena Mangano, Ashley Prince, Alex Feliz, Merri Rosen, Amy Wolf, Pam Speigel, Maria Poveromo, Peter Kim, Glenda Insabella, Cindy Nye, Suzanne Margiano, Diane Schrage, and Jean Silver.

### Notes

1. This is not meant to imply that infants initially have universal abilities to perceive all contrasts from all languages, although this has sometimes been claimed or implied. Certain contrasts, both native and nonnative, may be more difficult than others for young infants to discriminate (e.g., Aslin et al. 1981; Eilers and Minifie 1975; Eilers et al. 1981; Eilers, Wilson, and Moore 1979; Kuhl 1980; see further discussion in section on *Language-particular developments in infant speech perception*). The point here is that initially infants' discrimination of segmental contrasts does not yet appear to be constrained by the particular language environment.
2. One alternative is that the auditory system begins and remains physiologically incapable of registering certain acoustic properties of speech unless the listeners' environment provides exposure to those properties, which would then induce sensitivity (presumably during some critical developmental period). This would be a strictly sensory-neural version of the induction hypothesis formulated by Aslin and Pisoni (1980a; see also Pisoni, Lively, and Logan this volume) to explain one possible form of experiential effect on perceptual development, which was based on Gottlieb's (1981) model of visual and auditory development in ducklings. Another alternative effect of auditory experience is that some phonetic contrasts are weak in acoustic salience and hence initially difficult for the infant to discriminate but that relatively frequent exposure to these can improve discrimination (Eilers and Oller 1988; Eilers, Wilson, Moore 1979).

The logical problem with the absolute notion of sensory-neural induction is that, if the sensory system is incapable of registering an acoustic property, the exposure is intrinsically ineffectual and incapable of inducing sensitivity. To illustrate, the human visual system cannot register ultraviolet wavelengths as visible light, and our abundant exposure to ultraviolet light from both natural and artificial sources never induces visual sensitivity to those wavelengths.

The acquisition of perceptual ability to discover previously unrecognized organization in stimulation is, however, another matter so long as the sensory system is already capable of registering the supporting physical evidence. The emergence of such perceptual abilities certainly does occur associated with adjustments of selective attention (Gibson 1966; Gibson and Gibson 1955). In these cases, the term perceptual learning is preferable to induction. Sensory capacity is necessary, though not sufficient, for perceptual learning, but perceptual learning cannot induce sensory capacity. In other words, if the system cannot register a stimulus property, experience will not change that fact. If the system can register the property, but the perceiver does not initially recognize the pattern of information it conveys, then experience can lead to perceptual learning. Confusion about induction may arise within a theoretical model to the extent that it conflates experiential effects on physiological mechanisms with experiential effects on perceptual skills (i.e., selective attention).

3. Space limitations prevent a full recapitulation of the discussion here; for additional details, the reader is referred to the original papers.

4. A very similar question about adult perception and production of nonnative sounds has been addressed in the work of Flege (1988, Flege and Eefting 1986, 1987), whose discussion of the issue is compatible in some respects with the view presented here but divergent in other.

5. The diversity of possible assimilations is due to the fact that Zulu /ɬ/ shares partial articulatory commonalities with each of these (and perhaps other) English consonants and is simultaneously discrepant from each of them on other articulatory properties, that is, the gestural coordination is unfamiliar with respect to English. As stated earlier, listeners should show sensitivity to both similarities and discrepancies between the nonnative phone and native categories. Furthermore, they may vary regarding which similarities/discrepancies capture their attention; hence they may differ regarding which of several possible native categories assimilates a given nonnative phone.

## References

- Aitchinson, J. (1983) *The articulate mammal*. New York: Universe Press.
- Aslin, R. N. and Pisoni, D. B. (1980a). Effects of early linguistic experience on speech discrimination by infants: A critique of Eilers, Gavin, and Wilson (1979). *Child Development*, 51, 107-112.
- Aslin, R. N. and Pisoni, D. B. (1980b). Some developmental processes in speech perception. In G. Yeni-Komshian, J. F. Kavanaugh, and C. A. Ferguson (eds.), *Child phonology*, vol 1: *Production*. New York: Academic Press.
- Aslin, R. N., Pisoni, D. B., Hennessy, B. L., and Perey, A. J. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. *Child Development*, 52, 1135-1145.
- Aslin, R. N., Pisoni, D. B., and Jusczyk, P. W. (1983). Auditory development and speech perception in infancy. In M. M. Haith and J. J. Campos (eds.), *Handbook*

of child psychology, vol. 2: *Infancy and developmental psychobiology* (pp. 573-687). New York: Wiley.

Bahrick, L. E. (1987). Infants' intermodal perception of two levels of temporal structure in natural events. *Infant Behavior and Development*, 10, 387-416.

Bahrick, L. E. (1988). Intermodal learning in infancy: Learning on the basis of two kinds of invariant relations in audible and visible events. *Child Development*, 59, 197-209.

Best, C. T. (1984). Discovering messages in the medium: Speech and the pre-linguistic infant. In H. E. Fitzgerald, B. Lester, and M. Yogman (eds.), *Advances in pediatric psychology*, vol. 2 (pp. 97-145). New York: Plenum Press.

Best, C. T. (1990). Adult perception of nonnative contrasts differing in assimilation to native phonological categories. *Journal of the Acoustical Society of America*, 88, S177.

Best, C. T. (1991a). Language-specific effects on perception of discourse prosody categories by 2-4 month olds. Presented at the International Conference on Event Perception and Action, Amsterdam, The Netherlands, August 1991.

Best, C. T. (1991b). Phonetic influences on the perception of nonnative speech contrasts by 6-8 and 10-12 month-olds. Presented at the meeting of the Society for Research in Child Development, Seattle, Wash., April 1991.

Best, C. T. and McRoberts, G. (1989). Phonological influences on infants' perception of two nonnative speech contrasts. Paper presented to the Society for Research in Child Development, Kansas City, Mo, April 1989.

Best, C. T., Goodell, E., and Wilkenfeld, D. (in preparation). Phonologically-motivated substitutions in two- and five-year-olds' imitations of intervocalic alveolar stops.

Best, C. T., Levitt, A. G., and McRoberts, G. W. (1991). Examination of language-specific influences in infants' discrimination of prosodic categories. *Proceedings of the 12th international congress of phonetic sciences*. Aix-en-Provence, France.

Best, C. T., McRoberts, G. W., Goodell, E., Womer, J. S., Insabella, G., Kim, P., Klatt, L., Luke, S., and Silver, J. (1990). Infant and adult perception of nonnative speech contrasts differing in relation to the listener's native phonology. *Infant Behavior and Development*, 13 (Abstract).

Best, C. T., McRoberts, G. W., and Sithole, N. N. (1988). The phonological basis of perceptual loss for nonnative contrasts: Maintenance of discrimination among Zulu clicks by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 345-360.

Best, C. T., Morrongiello, B., and Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception and Psychophysics*, 29, 191-211.

- Brière, E. (1966). An investigation of phonological interference. *Language*, 42, 769-796.
- Browman, C. P. and Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, 3, 219-252.
- Browman, C. P. and Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, 62, 201-251.
- Browman, C. P. and Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49, 155-180.
- Burnham, D. K. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics*, 7, 207-240.
- Carney, A. E., Widin, G. P., and Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, 62, 961-970.
- Dent, C. (1990). An ecological approach to language development: An alternative to functionalism. *Developmental Psychobiology*, 23, 679-703.
- Diehl, R. and Kluender, K. (1989). On the objects of speech perception. *Ecological Psychology*, 1, 121-144.
- Eilers, R. E. (1977). Context-sensitive perception of naturally produced stop and fricative consonants by infants. *Journal of the Acoustical Society of America*, 61, 1321-1336.
- Eilers, R. E. and Minifie, F. D. (1975). Fricative discrimination in early infancy. *Journal of Speech and Hearing Research*, 18, 158-167.
- Eilers, R. E. and Oller, D. K. (1988). Precursors to speech: What is innate and what is acquired? *Annals of Child Development*, 5, 1-32.
- Eilers, R. E., Gavin W. J., and Oller, D.K. (1982). Cross linguistic perception in infancy: The role of linguistic experience. *Journal of Child Language*, 9, 289-302.
- Eilers, R. E., Gavin W. J., and Wilson, W. R. (1979). Linguistic experience and phonemic perception in infancy: A cross-linguistic study. *Child Development*, 50, 14-18.
- Eilers, R. E., Gavin W. J., and Wilson, W. R. (1980). Effects of early linguistic experience on speech discrimination by infants: A reply. *Child Development*, 51, 113-117.
- Eilers, R. E., Morse, P. A., Gavin, W. J., and Oller, D. K. (1981). The perception of voice-onset-time in infancy. *Journal of the Acoustical Society of America*, 70, 955-965.
- Eilers, R. E., Oller, D. K., Bull, D. H., and Gavin, W. J. (1984). Linguistic experience and infant speech perception: A reply to Jusczyk, Shea and Aslin (1984). *Journal of Child Language*, 11, 467-475.



Eilers, R. E., Wilson, W. R., and Moore, J. M. (1977). Developmental changes in speech discrimination in infants. *Journal of Speech and Hearing Research*, 20, 766-780.

Eilers, R. E., Wilson, W. R., and Moore, J. M. (1979). Speech discrimination in the language-innocent and the language-wise: A study in the perception of voice onset time. *Journal of Child Language*, 6, 1-18.

Eimas, P. D. (1974). Linguistic processing of speech by infants. in R. L. Schiefelbusch and L. L. Lloyd (eds.), *Language perspectives—Acquisition, retardation, and intervention*. Baltimore: University Park Press.

Eimas, P. D. (1975). Speech perception in early infancy. In L. B. Cohen and P. Salapatek (eds.), *Infant perception: From sensation to cognition*. New York: Academic Press.

Eimas, P. D. (1978). Developmental aspects of speech perception. in R. Held, H. Leibowitz, and H. L. Teuber (eds.), *Handbook of sensory physiology*, vol. 8: *Perception*. New York: Springer-Verlag.

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

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Ferguson, C. A. (1986). Discovering sound units and constructing sound systems: It's child's play. In J. S. Perkell and D. H. Klatt (eds.), *Invariance and variability of speech processes* (pp. 36-53). Hillsdale, N.J.: Erlbaum.

Flege, J. E. (1984). The detection of French accent by American listeners. *Journal of the Acoustical Society of America*, 76, 692-707.

Flege, J. E. (1987). A critical period for learning to pronounce foreign languages? *Applied Psycholinguistics*, 8, 162-177.

Flege, J. E. (1988). The production and perception of speech sounds in a foreign language. In H. Winitz (ed.), *Human communication and its disorders: A review 1988*. Norwood, N.J.: Ablex.

Flege, J. E. (1990). Perception and production: The relevance of phonetic input to L2 phonological learning. In C. Ferguson and T. Huebner (eds.), *Crosscurrents in second language acquisition and linguistic theories*. Philadelphia: John Benjamins.

Flege, J. E. and Eefting, W. (1986). Linguistic and developmental effects on stop-perception and production by native speakers of English and Spanish. *Phonetica*, 43, 323-347.

Flege, J. E. and Eefting, W. (1987). The production and perception of English stops by Spanish speakers of English. *Journal of Phonetics*, 15, 67-83.

Flege, J. E., McCutcheon, M., and Smith, S. (1987). The development of skill in producing word-final English stops. *Journal of the Acoustical Society of America*, 82, 433-447.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, Mass.: MIT Press.

- Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14, 3-28.
- Fowler, C. A. (1989). Real objects of speech perception: A commentary on Diehl and Kluender. *Ecological Psychology*, 1, 145-160.
- Fowler, C. A. (1991). Sound-producing sources as objects of perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America*, 88, 1236-1249.
- Fowler, C. A. and Dekle, D. J. (1991). Listening with eye and hand: Cross-Modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 816-828.
- Fowler, C. A. and Smith, M. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. Perkell and D. H. Klatt (eds.), *Invariance and variability in speech processes* (pp. 123-139). Hillsdale, N.J.: Erlbaum.
- Fowler, C. A. and Turvey, M. T. (1978). Skill acquisition: An event approach with special reference to searching for the optimum of a function of several variables. In G. Stelmach (ed.), *Information processing in motor control and learning*. New York: Academic Press.
- Fowler, C. A., Best, C. T., and McRoberts, G. W. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception and Psychophysics*, 48, 559-570.
- Gardner, B. T. and Gardner, R. A. (1973). Evidence for sentence constituents in the early utterances of child and chimpanzee. *Journal of Experimental Psychology: General*, 104, 244-267.
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Gibson, J. J. and Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review*, 62, 32-41.
- Gillette, S. (1980). Contextual variation in the perception of L and R by Japanese and Korean speakers. *Minnesota Papers in Linguistics and the Philosophy of Language*, 6, 59-72.
- Goodell, E. and Studdert-Kennedy, M. (1990). From phonemes to words or words to phonemes: How do children learn to talk? Paper presented at the International Conference on Infant Studies, Montreal, April 1990.
- Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R." *Neuropsychologia*, 9, 317-323.
- Gottlieb, G. (1981). The roles of early experience in species-specific perceptual development. In R. N. Aslin, J. R. Alberts, and M. R. Peterson (eds.), *Develop-*

*ment of perception: Psychobiological perspectives*, vol. 1: *Audition, somatic perception, and the chemical senses*. New York: Academic Press.

Grieser, D. A. and Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology*, 25, 577-588.

Insabella, G. (1990). *Four-year-olds' perception of nonnative phonetic contrasts*. Unpublished senior honors thesis, Wesleyan University, Middletown, Conn.

Insabella, G. and Best, C. T. (1990). Four-year-olds' perception of nonnative contrasts differing in phonological assimilation. Paper presented at meeting of the Acoustical Society of America, San Diego, November 1990.

Jusczyk, P. W. (1981). Infant speech perception: A critical appraisal. In P. D. Eimas and J. A. Miller (eds.), *Perspectives in the study of speech*. Hillsdale, N.J.: Erlbaum.

Jusczyk, P. W. (1985). On characterizing the development of speech perception. In J. Mehler and R. Fox (eds.), *Neonate cognition: Beyond the blooming, buzzing confusion* (pp. 199-229). Hillsdale, N.J.: Erlbaum.

Jusczyk, P. W. (1986). Toward a model of the development of speech perception. In J. S. Perkell and D. H. Klatt (eds.), *Invariance and variability in speech processes*. Hillsdale, N.J.: Erlbaum.

Jusczyk, P. W., Shea, S. L., and Aslin, R. N. (1984). Linguistic experience and infant speech perception: A re-examination of Eilers, Gavin and Oller (1982). *Journal of Child Language*, 11, 453-466.

Kewley-Port, D. and Preston, M. S. (1974). Early apical stop production: A voice onset time analysis. *Journal of Phonetics*, 2, 195-210.

Kuhl, P. K. (1980). Perceptual constancy for speech-sound categories in early infancy. In G. H. Yeni-Komshian, J. F. Kavanaugh, and C. A. Ferguson (eds.), *Child phonology*, vol. 2: *Perception* (pp. 41-66). New York: Academic Press.

Kuhl, P. K. (1987). Perception of speech and sound in early infancy. In P. Salapatek and L. Cohen (eds.), *Handbook of infant perception*, vol. 2 (pp. 275-382). New York: Academic Press.

Kuhl, P. K. (1991). 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.

Kuhl, P. K. and Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, 218, 1138-1144.

Kuhl, P. K. and Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior and Development*, 7, 361-381.

Ladefoged, P. (1981). *A course in phonetics* (2nd ed.) New York: Harcourt, Brace, and Jovanovich.

- Lalonde, C. and Werker, J. (1990). Cognitive/perceptual integration of three skills at 9 months. Paper presented at the International Conference on Infant Studies, Montreal, April 1990.
- Lasky, R. E., Syrdal-Lasky, A., and Klein, R. E. (1975). VOT discrimination by 4 to 6½ month old infants from Spanish environments. *Journal of Experimental Child Psychology*, 20, 215-225.
- Leonard, L. B., Fey, M. E., and Newhoff, M. (1981). Phonological considerations in children's early imitative and spontaneous speech. *Journal of Psycholinguistic Research*, 10, 123-133.
- Leonard, L. B., Schwartz, R. G., Folger, M. K., and Wilcox, M. J. (1978). Some aspects of child phonology in imitative and spontaneous speech. *Journal of Child Language*, 5, 403-415.
- Levitt, A., Jusczyk, P., Murray, J., and Carden, G. (1988). Context effects in two-month-old infants' perception of labiodental/interdental fricative contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 361-368.
- Liberman, A. M. and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1-36.
- Liberman, A. M. and Mattingly, I. G. (1989). A specialization for speech perception. *Science*, 245, 489-494.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431-461.
- Lieberman, P. (1975). *On the origins of language: An introduction to the evolution of human speech*. New York: Macmillan.
- Lindblom, B., MacNeilage, P., and Studdert-Kennedy, M. (1983). Self-organizing processes and the explanation of phonological universals. In B. Butterworth, B. Comrie, and O. Dahl (eds.), *Universals workshop*. The Hague: Mouton.
- Lisker, L. and Abramson, A. S. (1970). The voicing dimension: Some experiments on comparative phonetics. *Proceedings of the 6th international congress of phonetic sciences, Prague, 1967* (pp. 563-567). Prague: Academia.
- Locke, J. L. (1983). *Phonological acquisition and change*. New York: Academic Press.
- MacDonald, J. and McGurk, H. (1978). Visual influences on speech perception processes. *Perception and Psychophysics*, 24, 253-257.
- McClasky, C. L., Pisoni, D. B., and Carrell, T. D. (1983). Effects of transfer of training on identification of a new linguistic contrast in voicing. *Research on Speech Perception Progress Report (Indiana University)*, 6, 205-234.
- McCune, L. and Vihman, M. (1987). Vocal motor schemes. *Papers and Reports in Child Language Development*, 26.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.

MacKain, K.S. (1982). Assessing the role of experience on infants' speech discrimination. *Journal of Child language*, 9, 527-542.

MacKain, K. S., Best, C. T., and Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, 2, 369-390.

MacKain, K. S., Studdert-Kennedy, M., Spieker, S., and Stern, D. (1983). Infant intermodal speech perception is a left hemisphere function. *Science*, 219, 1347-1349.

Macken, M.A. (1979). Developmental reorganization of phonology: A hierarchy of basic units of acquisition. *Lingua*, 49, 11-49.

Macken, M. A. and Ferguson, C. A. (1983). Cognitive aspects of phonological development: Model, evidence, and issues. In K. E. Nelson (ed.), *Children's language*. Hillsdale, N.J.: Erlbaum.

Mann, V. A. (1980). Influence of preceding liquid on stop consonant perception. *Perception and Psychophysics*, 28, 407-412.

Mann, V. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of "l" and "r." *Cognition*, 24, 169-196.

Massaro, D. M. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, N.J.: Erlbaum.

Mattingly, I. G. and Liberman, A. M. (1988). Specialized perceiving systems for speech and other biologically significant sounds. In G. M. Edelman, W. E. Gall, and W. M. Cowan (eds.) *Auditors function: Neurobiological bases of hearing* (pp. 775-793). New York: Wiley.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoni, J., and Amiel-Tison, C. A. (1988). A precursor of language acquisition in young infants. *Cognition*, 29, 143-178.

Menn, L. (1971). Phonotactic rules in beginning speech. *Lingua*, 26, 225-251.

Menn, L. (1978). Phonological units in beginning speech. In A. Bell and J. Hooper (eds.), *Syllables and segments*. Amsterdam: North-Holland.

Menn, L. (1986). Language acquisition, aphasia and phonotactic universals. In F. R. Eckman, E. A. Moravcsik, and J. R. Wirth (eds.), *Markedness*. (pp. 241-255). New York: Plenum Press.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., and Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics*, 18, 331-340.

Mochizuki, M. (1981). The identification of /r/ and /l/ in natural and synthesized speech. *Journal of Phonetics*, 9, 283-303.

- Moskowitz, B. A. (1980). Idioms in phonology acquisition and phonological change. *Journal of Phonetics*, 8, 69-83.
- Nittrouer, S., Studdert-Kennedy, M., and McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research*, 32(1), 120-132.
- Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. Yeni-Komshian, J. F. Kavanaugh, and C. A. Ferguson (eds.), *Child phonology*, vol. 1: *Production*. New York: Academic Press.
- Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, 5, 261-283.
- Pisoni, D. B. and Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America*, 55, 328-333.
- Pisoni, D. B., Aslin, R. N., Perey, A. J., and Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 297-314.
- Premack, D. (1971). Language in chimpanzee? *Science*, 172, 808-822.
- Rosenblum, L. (1987). Towards an ecological alternative to the motor theory of speech perception. *Perceiving-Acting Workshop Review (Tech. Rep. CESPA, University of Connecticut)*, 2, 25-28.
- Sheldon, A. and Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, 3, 243-261.
- Stark, R. E. (1980). Stages of speech development in the first year of life. In G. H. Yeni-Komshian, J. F. Kavanaugh, and C. A. Ferguson (eds.), *Child phonology*, vol. 1: *Production*. New York: Academic Press.
- Strange, W. and Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception and Psychophysics*, 36, 131-145.
- Streeter, L. A. (1976a). Kikuyu labial and apical stop discrimination. *Journal of Phonetics*, 4, 43-49.
- Streeter, L. A. (1976b). Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature*, 259, 39-41.
- Studdert-Kennedy, M. (1981). The emergence of phonetic structure. *Cognition*, 10, 301-306.
- Studdert-Kennedy, M. (1985). Perceiving phonetic events. In W. H. Warren and R. E. Shaw (eds.), *Persistence and change*. Hillsdale, N.J.: Erlbaum.

Studdert-Kennedy, M. (1986a). Development of the speech perceptuomotor system. In B. Lindblom and R. Zetterstrom (eds.), *Precursors of early speech*. New York: Stockton Press.

Studdert-Kennedy, M. (1986b). Sources of variability in early speech development. In J. S. Perkell and D. H. Klatt (eds.), *Invariance and variability in speech processes*. Hillsdale, N.J.: Erlbaum.

Studdert-Kennedy, M. (1987). The phoneme as a perceptuomotor structure. In A. Allport, D. MacKay, W. Prinz, and E. Scheerer (eds.), *Language perception and production* (pp. 67-84). London: Academic Press.

Studdert-Kennedy, M. (1989). The early development of phonological form. In C. von Euler, H. Forssberg, and H. Lagercrantz (eds.), *Neurobiology of early infant behavior* (pp. 287-301). Basingstoke, England: Macmillan.

Studdert-Kennedy, M. (1991). Language development from an evolutionary perspective. In N. Krasnegor, D. Rumbaugh, R. Scheifelbusch, and M. Studdert-Kennedy (eds.), *Language acquisition: Biological and behavioral determinants*. Hillsdale, N.J.: Erlbaum.

Tahta, S., Wood, M., and Loewenthal, K. (1981). Foreign accents: Factors relating to transfer of accent from the first language to a second language. *Language and Speech*, 24, 265-272.

Tees, R. C. and Werker, J. F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate nonnative speech sounds. *Canadian Journal of Psychology*, 38, 579-590.

Terrace, H. S., Petitto, L. A., Sanders, R. J., and Bever, T. G. (1979). Can an ape create a sentence? *Science*, 206, 809-902.

Trehub, S. E. (1973). Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology*, 9, 91-96.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by adults and infants. *Child Development*, 47, 466-472.

Waterson, N. (1971). Child phonology: A prosodic view. *Journal of Linguistics*, 7, 179-211.

Werker, J. F. (1989). Becoming a native listener. *American Scientist*, 77, 54-59.

Werker, J. F. and Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, 24, 672-683.

Werker, J. and Logan, J. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics*, 37, 35-44.

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

Werker, J. F. and Tees, R. C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, 75, 1866-1878.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., and Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, 52, 349-355.