946.

ADVANCES IN INFANCY RESEARCH

VOLUME 9

1994

Co-editors

Carolyn Rovee-Collier
Department of Psychology
Rutgers University
New Brunswick, New Jersey

Lewis P. Lipsitt
Department of Psychology
and Child Study Center
Brown University

Valenza, E., Simion, F., & Umilta, C. (in press). Inhibition of return in newborns. *Infant Behavior and Development.*
Van Hof-van Duin, J., & Mohn, G. (1986). Visual field measurements, optokinetic nystagmus, and the threatening response: Normal and abnormal development. *Documenta Ophthalmologica Proceedings Series, 45,* 305–315.
Weiskrantz, L. (1986). *Blindsight: A case study and implications.* Oxford: Clarendon Press.
Yuodelis, C., & Hendrickson, A. (1986). A qualitative and quantitative analysis of the human fovea during development. *Vision Research, 26,* 847–855.

LEARNING TO PERCEIVE THE SOUND PATTERN OF ENGLISH*

Catherine T. Best

WESLEYAN UNIVERSITY AND HASKINS LABORATORIES

*Preparation of this chapter was supported by an NIH Research Career Development award from the National Institute on Deafness and Communication Disorders (DC00045). Research from my lab that is described herein was also supported by NIH research grant DC00403 to the author and by program project grant HD01994 to Haskins Laboratories. Many thanks are due to Alice Faber, Andrea Levitt, Steven Braddon, Janet Werker, and Peter Jusczyk for their helpful comments on an earlier draft of the chapter; to Alice also for managing the infant lab in my near absence; and to a wonderful group of research assistants who kept the lab running smoothly while I wrote: Janet Calderòn, Sandy Chiang, Ron Dewitt, Domenica Giancolo, and Alyssa Wulf. Thanks also to Steve for doing more than his fair share of child care, and to my daughters, Aurora and Vanessa, for being patient.

I. INTRODUCTION

Language lies at the heart of human cognitive and social development. Infants, who are by definition "without language," become speaker–hearers of particular languages within their first few years through their experience with the speech of their caregivers and other significant people in their environment. The foundation for the emergence of language proper is the infant's discovery of sound–meaning correspondences in the utterances produced by those significant people. Social and physical contexts provide support for the semantic meaning of an utterance, although determining the specific referent of an unknown word from nonlinguistic context alone may be no simple task (see Quine, 1960). The present discussion, however, focuses on the other side of the sound–meaning relation—the sound pattern itself. It is still far from clear how the infant comes to recognize in the stream of connected speech the sequence of consonants and vowels that may underlie the diverse pronunciations of a given word in different sentences, by different speakers, and under different speaking conditions (e.g., in rapid casual speech versus slow, exaggerated infant-directed speech). Presumably, these accomplishments are built on the infant's prior abilities to discriminate and classify the audible properties that correspond to various levels of organization in speech, for example, consonants and vowels (phonetic segments), rhythmic stress patterns, prosodic phrases, and so forth.

It is these perceptual abilities for handling the "surface phonetic structure" of speech that are the primary concern of this chapter. In particular, we focus on how the infant's experience with a particular language begins to influence perception of consonant and vowel contrasts that fall outside the phonetic inventory employed by that language. Developmental changes in perception of such nonnative contrasts can provide important insights about the aspects of the native phonological system to which infants are becoming attuned as they gain experience with native speech. The central goal of this chapter is to describe and provide evidence for a model of how language-specific experience influences infants' and adults' perception of nonnative phonetic contrasts. The model is the Perceptual Assimilation Model of cross-language speech perception.

First, however, we briefly review the basic pattern of developmental change in perception of nonnative phonetic contrasts and describe the phonetic and phonological organization in spoken language that the infant must come to perceive. Following that introduction to speech and its perceptual requirements, we consider two major theoretical perspectives that might be extended to account for language-specific developmental changes to provide a backdrop for the presentation of the Perceptual Assimilation Model.

II. INFANTS' PERCEPTION OF PHONETIC PROPERTIES IN SPEECH

Young infants can discriminate a wide range of phonetic contrasts between consonants (e.g., [b] vs. [d]) or between vowels (e.g. the vowels in *boot* vs. *book*), whether or not the tested phonetic features are employed linguistically by the ambient language. But by adulthood, in fact by much earlier in development, experience with the native language comes to exert some rather striking effects on the perception of phonetic contrasts. The experiential influence is particularly apparent for perception of contrasts that are not part of the native language's phonological system. As explained more fully in the next section, the phonological system refers to the rules by which a given language employs certain phonetic differences as linguistic contrasts that can convey differences in word meanings. It treats certain other phonetic differences as linguistically equivalent, and yet other phonetic features as nonpermissible altogether, even though the same features may be used linguistically by some other language. Mature listeners often have substantial difficulty discriminating and categorizing phonetic contrasts that are not part of their own phonological system, but young infants from the same language environment have no difficulty discriminating those same contrasts. Effects of language-specific experience emerge in speech perception during the second half of the infant's first year and are clearly evident by 10–12 months for perception of many nonnative consonant contrasts (see reviews by Best, 1984, 1993, in press-a; Werker, 1989, 1991; Werker & Pegg, 1992).

Why and how does experience with the native language come to shape the perception of the phonetic properties of speech in this manner? How do infants become familiar with the sound system of their native language, and how does that process subsequently shape perception of unfamiliar consonants and vowels from languages not heard before? Infants' initial experience with their language begins with only the surface phonetic patterns of spoken utterances, but ultimately they must use that input to develop knowledge of the underlying semantic concepts and syntactic rules of the language. Thus, the first inroads the infant makes into discovering the systematic structure of the language take place at some level of its sound system. Many believe that this discovery process commences at the prosodic level.

Recent research on prosodic bootstrapping—the notion that conversational speech (particularly infant-directed speech) provides converging intonational and rhythmic markers that guide infants' attention to clause and phrase boundaries in speech—has made important advances in our understanding of how infants may discover the boundaries of syntactic units at varying levels (e.g., Gleitman, Gleitman, Landau, & Wanner, 1988; Hirsch-Pasek et al., 1987; Jusczyk & Kemler Nelson, in press; Kemler Nelson, Hirsch-Pasek, Jusczyk, & Wright-Cassidy, 1989; Morgan, 1990). However, prosodic bootstrapping may not help the infant so much with segmenting sound at the word level. Broad prosodic markers do not consistently specify word boundaries in continuous speech (cf. Gerken & McIntosh, 1993; Jusczyk, Cutler, & Redanz, 1993), especially in languages like French, which lacks syllabic stress alternation patterns like those found in English. But word boundaries are often marked by characteristic differences in the exact way that the surrounding consonants and/or vowels are pronounced (e.g., aspirated [t] and reduced "uh" vowel in *citrus* but not in *sit Russ*), phonetic characteristics to which even very young infants appear to be sensitive (Christophe, Dupoux, Bertoncini, & Mehler, submitted; Hohne & Jusczyk, 1992). Thus, word segmentation may be aided not so much (or not only) by prosodic bootstrapping, but more by what might be called *phonetic bootstrapping.*

It is the infant's attention to this sort of detailed phonetic information that would seem to be most revelant to the discussion of how language-specific experience begins to influence perception of consonants and vowels, also referred to as *phonetic segments.* A basic premise of this chapter is that infants make use of surface phonetic details to discover the more abstract phonological properties of their native language. As described more fully in a subsequent section, the phonological system refers to the inventory of phonetic segments that a given language employs to convey meaningful differences among words. This inventory is organized systematically and hierarchically around multiple contrasting phonetic features that define linguistically important relations among phonetic segments. The systematicity of a language's phonological system makes possible the vast expansion of vocabulary that takes place in early childhood and serves somewhat later as the linguistic framework for the child's acquisition of reading and writing abilities. But the relation between the surface phonetic details of utterances and the more abstract phonological system of a language is not always transparent, in part because of contextually determined differences in the phonetic details of consonants and vowels and other effects such as speaker and speaking-rate differences in pronunciations. Thus, in order to learn the sound pattern of the ambient language sufficiently to determine sound-meaning relations, the infant must begin to untangle the complex relationship between the surface phonetics and the underlying phonological system, at least to some approximation.

To provide a foundation for considering developmental changes in speech perception, we turn now to an overview of the hierarchical nesting of linguistic information conveyed in the speech signal. We focus in particular on the relationship between the lower order patterning at the surface phonetic level of speech and the more abstract, higher order organization at the phonological level of a given language. Differences in the sound patterns of different languages reflect differences not only in their inventories of consonants and vowels, but also especially in the patterns by which they relate phonetic details to phonological structure. It is the relationship between phonetic details and phonological organization that is most germane to understanding the effects of language experience on the perception of nonnative speech-sound contrasts. Any theory of the acquisition of native language sound patterns, and of the perception of those patterns, must be able to take into account the sound structure of the spoken message and the observations of language- and dialect-specific differences in that structure.

III. THE STRUCTURE OF THE SPOKEN MESSAGE

When we convey a spoken message to a listener, the utterance we produce via the audible, and to some extent visible, articulatory movements of our vocal tract is organized according to the multiple levels of linguistic structure of the language we speak (the property of dual structure; Hockett, 1963). That is, the spoken utterance concurrently reflects the organizing of sound into words, the syntactic organization of those words into the larger units of noun, verb, or other phrases, and the superordinate syntactic organization of phrases into clauses, one or more of which may comprise a sentence. At the same time, prosodic organization is evident in the intonation, temporal patterns, and amplitude changes that provide a common carrier for the words at the phrase, clause, and sentence levels and serve to signal linguistic stress, pragmatic emphasis, and emotional tone. But there is also nested structure if we look in the opposite direction, below the level of individual words. A word is composed of one or more units of meaning, referred to as *morphemes;* for example, the word *incomplete* contains the stem morpheme *complete* plus the negation prefix *in-*. Morphemes are comprised of one or more syllables, each made up of consonants and vowels, which are defined in standard linguistic analysis as *phonological segments.*

A. Phonological Patterning

Phonological segments are the smallest units of the language-specific grammatical system. They are themselves composed of *phonetic features,* the matrix of articulatory/acoustic properties that characterize the way a given phoneme is produced. These properties are described according to a universal set of distinctive feature contrasts by which one segment can differ critically from all others (e.g., Jakobson, Fant, & Halle, 1963; for an introduction to phonetics, see Catford, 1988; Ladefoged, 1982). For example, the consonants and vowels in the word *incomplete* may be broadly transcribed to correspond to phonemic segments as /ɪnkamplit/. However, additional phonetic details that are present in the actual production of the word can be represented in a narrow phonetic transcription as [ɪŋkəmp^{h}l̥it̚]. The narrow transcription indicates that the /n/ preceding the /k/ is actually produced as a nasalized constriction [ŋ] near the soft palate at the back of the mouth, rather at the alveolar ridge behind the upper front teeth [n]. The vowel in the second, unstressed syllable is the reduced vowel schwa [ə], which is somewhat like the "uh" ([ʌ]) in *butter,* but shorter in duration. The /p/ is produced with breathy aspiration [p^h], which causes the following /l/ to be devoiced [l̥]. And the tongue-tip closure for the final /t/ is not audibly released at the end of the word [t̚]. (For an introduction to phonology, see Kenstowicz & Kisseberth, 1979.)

The phonology of a language is the set of systematic constraints the language places on the sound patterning of its consonants and vowels. To begin with, every language employs but a subset of all humanly producible consonant and vowel sounds to produce minimal phonological contrasts in word meanings. As an illustration of minimal contrast, English uses /b/ and /p/ to differentiate the meaning of words that are matched in their other phonemic elements, such as *bat* versus *pat.* Likewise, the vowel contrast /ɪ/–/ɛ/ distinguishes the minimally contrasting words *pit–pet* (/pɪt/–/pɛt/). However, modern English lacks the throaty fricative at the beginning of the Yiddish word *chutzpah.*

The phonology of a language also includes contextually determined allophonic variations in the phonetic details of a given phoneme produced in different surrounding contexts. For example, in English the /p/ in *pan* is produced with aspiration and a long lag before voicing starts after the release of the bilabial closure, denoted phonetically as [p^h]. But the /p/ in *span* is produced with a much shorter voicing lag and without aspiration, denoted as the allophone [p]. However, this difference in pronunciation does not signal a phonological contrast in English. Phonological analyses of the range and constraints on allophonic variants reveal which one is the underlying phonological form and which others are the variants of that underlying form. In this case, [p] is a variant of underlying [p^h]. There are no English minimal word pairs whose meaning is differentiated phonologically solely by the /p/–/p^h/ difference.

Certain other contextually determined effects on the phonetic details of segments in a spoken message result from more global changes, such as different speech rates and styles. To illustrate, the phrase *did you eat . . .* in slow, careful speech is typically produced with two clear /d/'s and the "ih" vowel in *did,* clear "y" and long "oo" sounds for *you,* and a clear "ee" and /t/ in *eat*. But in rapid, casual speech the phrase may become *d'y'eat . . . ,* where the initial /d/ and vowel in *did* have been omitted, the final /d/ seems to combine with the "y" of *you* to form a "j" sound, and the long "oo" has become an unstressed schwa [e] (e.g., Browman & Goldstein, 1990a; Oshika, Zue, Weeks, Neu, & Aurbach, 1975).

Languages also have phonotactic constraints on the distributional patterns of consonants and vowels, including permissible sequences in syllables and permissible positions that particular sounds can occupy within a syllable or word. For example, /spa/ and /mop/ (*mope*) are permissible English syllables, but */psa/ and */mpo/ are not. Also, English words may end but may not begin with the velar nasal /ŋ/ (as in *so<u>ng</u>*) or may have an internal voiced palatal fricative "zh" (as in *mea<u>s</u>ure*) but may not begin with this sound.

Thus, the phonological system of a language refers to the underlying linguistically defined relations among the consonant and vowel sounds it employs. The language's use of consonant or vowel differences for contrastive differentiation of word meanings, the allophonic patterning of those phonemes, and their phonotactic distributional constraints all reflect abstract invariant properties that underlie the surface phonetic details of spoken utterances. As should be clear from these examples, the relation between the phonetic details and the phonological organization of a language is often far from a simple, transparent mapping.

To address how infants might learn aspects of the language-specific phonology from ambient speech and how that might influence their perception of nonnative phonetic contrasts, we briefly review next how languages differ in the ways they relate the phonetic details of speech to phonological structure.

B. Language Differences in Phonology and Phonetics

An obvious way in which the sound patterns of languages differ is in their inventories of phonological segments and minimal contrasts. Although certain basic segment types seem to be universal, or nearly so, across the inventories of the world's languages, other sounds and contrasts are present only in some languages and are absent in others. Among the universally shared phonological segments are the stop consonants /p/ and /t/ and the vowels "ah" as in *father*, "ee" as in *see*, and "oo" as in *boot*.[1] Language differences in phonological inventories are numerous, however. For example, the /l/–/r/ contrast found in the inventory of English is absent from many Asian languages, such as Japanese and Korean, as well as from a number of other languages; indeed, the English /r/ is quite rare across languages. Similarly, the English vowels in *hook* and *hawk*, respectively, are lacking in Spanish, Native Hawaiian, and many other languages. Conversely, English lacks the click consonants of Zulu and other southern African languages, as well as the dental versus retroflex stop consonant contrast /d̪/–/ɖ/ of Hindi (our /d/ has a tongue-tip position between the Hindi sounds). English also lacks the front rounded vowels /y/–/ø/ found in French, German, and Swedish.

The neat and straightforward description of language differences in phonological inventories is seemingly complicated, however, by the fact that languages also either require or permit certain context-conditioned or free allophonic variants for at least some of their phonemes. For example, the French /r/ is characterized as a voiced uvular trill at the back of the throat, yet context conditioning causes its surface phonetic form to become a voiceless uvular fricative when it follows a voiceless consonant, for example, as in *quatre*, the French word for "four." Permissible differences among speakers also result in other freely varying allophones.

Allophonic variations may even, at times, appear to obfuscate claims that one language lacks a particular phoneme or contrast found in another. To illustrate, neither the dental nor the retroflex stop that contrast in Hindi are found in the English phonological inventory. Our /d/ is underlyingly a voiced alveolar stop [d]. However, a dental stop does occur phonetically in English speech, as an allophone of /d/ that is context conditioned due to coarticulation (overlapping production) with adjacent dental sounds. The dental allophone occurs when /d/ is adjacent to a dental fricative, for example, in *birthday*. These observations might seem to belie the claim that only Hindi, and not English, has a dental stop in its phonological inventory. The important point, though, is that this dental form does not contrast with /d/ in English. It is a context-conditioned allophone of /d/ and is heard as /d/. The adjacent dental segment is perceived as the source of the variant property (see also Fowler & Smith, 1986; Kent, Carney, & Severeid, 1974; Krakow, Beddor, Goldstein, & Fowler, 1988; Mann, 1980, 1986; Whalen,

---

[1]Exceptions are extremely rare. For example, Native Hawaiian lacks /t/, including instead only /p/ and /k/ for its nonnasal stop consonants.

1983), apparently even by young infants (Fowler, Best, & McRoberts, 1990).

The discussion about language differences in allophonic patterning prompts consideration of a similar phenomenon in which different languages, and different dialects of a single language, can differ in their phonetic realizations of the "same" phonological segment. If the phonetic details differ, then on what basis is the underlying segment in such cases "the same," in at least some crucial way? This question is more problematic for the cross-language case, but several observations suggest that underlying identify of segments, or at least close similarity, may often be a reasonable assumption nonetheless (see also Flege, 1987, in press). For one thing, the phonetic feature matrix that defines a given phonological segment includes only those features critical for distinguishing it from other segments in a language's phonology. Allophones are encompassed in the definition because they vary on noncritical features. Thus, English and Spanish both have the phonological segment /p/, even though it is often aspirated in English but never in Spanish. It is important to note, however, that listeners are quite sensitive to foreign accent in their native language, suggesting that listeners may nonetheless detect such subphonemic differences. Findings indicate that although some of the sensitivity to foreign accent is attributable to prosodic differences, for at least some cross-language segmental similarities, the phonetic differences between the corresponding native and nonnative segments are also perceptible (e.g., Flege, 1984, in press; Flege & Eefting, 1987; Flege & Fletcher, 1992).

Cross-language identity and similarity are corroborated by the phonological forms speakers use when learning a new language with unfamiliar pronunciations, as when a Spanish speaker's initial pronunciation of English *pit* may sound like *beet* because he or she uses the Spanish unaspirated /p/ and an "ee" vowel as Spanish has no "ih" sound. Cross-language segmental similarities are also suggested by the phonological forms speakers of one language give to loan words from another language (see also Silverman, 1992). For example, the French *calorique*, pronounced with an unaspirated /k/, an uvular trilled /r/, and the vowels "ah," "o," and "ee," has been adopted into English as *caloric* and pronounced with an English aspirated /k/, English /r/, and unstressed schwa [e] in the first and final vowel positions.[2] Moreover, similar sorts of phonological substitutions are seen in pidgins and creoles, interlanguages that result from social

[2]Although loan word pronunciations can be affected by spelling in both donor and recipient languages, the association between spelling and pronunciation is generally not arbitrary but reflects phonological principles. However, the degree of transparency between spelling and pronunciation differs among languages; for example, Spanish spelling is quite transparent, whereas English spelling is much less so.

contact between two independent language groups and that often derive only from spoken forms, at least in their early stages (e.g., Holm, 1988; Romaine, 1988). Finally, the patterns by which listeners label nonnative segments, not surprisingly, provide further converging evidence about cross-language segmental similarities, as described later.

By comparison to the cross-language case, the segmental identity issue seems relatively straightforward for the cross-dialect case, at first glance. For mutually intelligible dialects, the vocabulary, the grammar (phonology, morphology, syntax), and even the written forms are typically nearly identical between dialects. In this case, there is no doubt about phonological identity between corresponding segments in the dialects, even though they differ in some phonetic details. Here again, listeners nonetheless detect dialectal accent easily and show differential sensitivity to phonetic differences among segments in the native versus nonnative dialects (see Faber, Best, & Di Paolo, 1993).

Numerous examples of cross-dialect phonetic variants of underlying segments can be found in languages. On portions of Long Island in New York, words such as *long* are pronounced with a final /g/, although the final /g/ is omitted elsewhere in the United States. To take an example from another language, the nasalization of vowels in Canadian French commences later into the vowel than in continental French (van Reenen, 1982). Paralleling another between-language difference, one dialect may lack a phonological contrast found in other dialects of the same language (or found historically in the language), a situation termed a *merger* of the contrast. For example, English speakers from Canada, western United States, and areas of midwestern United States fail to produce or reliably label the "aw"–"ah" difference, as in *hawk–hock,* a vowel difference that is maintained in the northeastern United States (e.g., Di Paolo, 1992). Similarly, Texans have merged the "ih"–"eh" difference before /n/, pronouncing *pin* and *pen* as homonyms (both like *pin*).

Sometimes a merger is not absolute, but is rather a *near-merger* (see Faber, Di Paolo, & Best, submitted; Labov, 1974; Labov, Karen, & Miller, 1991). In a near-merger, a phonological contrast found elsewhere in the language is no longer evident in a given dialect, but productions of the near-merged sounds still show reliable acoustic differences and/or the contrast reappears in a subsequent sound change in the dialect. One such historical reversal occurred in early Modern English. The vowels in words like *meat–mate,* which had merged earlier, later reestablished different pronunciations when the *meat* class but not the *mate* class vowels merged with the vowel such as in words like *meet* (the *meat–meet* merger still stands today; Labov, 1974). As an example of a near-merger in current American English, /r/ is dropped after "ah" in some Boston dialects. Thus, word pairs such as *cod–card* are produced as near-homonyms (Costa & Mattingly, 1981). A similar effect is found in many dialects of British English. A near-opposite pattern occurs in Brooklyn, in which speakers add /r/-color to the "aw" sound, pronouncing *sauce* like *source* (Labov, Yaeger, & Steiner, 1972). In Albuquerque and the Salt Lake Valley, vowel pairs such as "ee"-"ih" and long "oo"-short "oo" (as in *boot-book*) show near-merger in the context of a following /l/. That is, word pairs such as *pool-pull* and/ *heel-hill* are pronounced as near-homophones (Di Paolo & Faber, 1991; Faber, 1992; Labov et al., 1972).

To return to cross-language differences, languages often differ in the phonotactic constraints they place on the sequences and word positions permitted among the segments in their inventories. As an illustration, English does not permit the "zh" sound word initially, but a number of other languages do, as in the French word for *magazine—journal*—and the Russian word for *woman—zhenshchina.* Likewise, English disallows "ng" ([ŋ]) in word-initial position, but that position is allowed in Vietnamese, such as in the name *Nguyen.* On the other hand, stop consonants such as /p/, /t/, and /k/ can occur in initial but not in final positions in Mandarin Chinese words and syllables; in English they can occur in either position. Finally, English phonotactics disallows certain phoneme sequences in syllables that are nonetheless permissible in other languages, such as */psa/ (e.g., in Greek), */mpo/ (e.g., in Chaga), and */dzva/ (e.g., in Polish).

In addition, the types of phonological alternations present in one language may be absent in others. As an example, Turkish uses a phonological principle of vowel-rounding harmony within words, whereby the vowels in a word must agree in whether they have lip rounding (e.g., "o" and long "oo") or not (e.g., "ee" or "ih"). Thus, the possessive form of *dere*—the word for *river*—is *deresi* but the possessive form of *boru*—the word for *pipe*—is *borusu.* English, of course, does not require any sort of vowel harmony. Other languages have a rule of vowel epenthesis to maintain a regular pattern of consonant-vowel alternation, whereby a vowel is inserted between any adjacent consonants. For example, pluralizing the Chuckchee word for *river—wejem*—by adding the plural morpheme *-ti* results in *wejemet* and not **wejemti* because the /m/ and /t/ must be separated by a vowel (the final *i* is deleted through a separate phonological rule). As a final example, some dialects of Spanish have a rule of spirantization by which voiced stop consonants /b/, /d/, and /g/ become voiced fricatives following a vowel, such as in the pronunciation of *nada*—the word for *no*—with a dental fricative instead of a /d/. It is interesting to note that the early words of young English-learning children often display phonological constraints that are absent from adult English but are similar to rules found in other languages. For example, complete vowel harmony is evident in *baba* for *bottle* and *dada* for *daddy,* whereas vowel epenthesis is evident in *buhlue* for *blue.* However, children's early phonologies sometimes also display other constraints that are seldom if ever seen in adult phonologies, such as the childish consonant harmony constraint by which *doggy* is produced as *dawddy* or *ducky* as *gucky*.

Language differences in phonological inventories and in the phonetic properties of identical or similar phonological segments are the primary aspects of phonology with which we deal in the remainder of the chapter. These are the aspects of speech most likely to be relevant to considering the lowest level invariants of native language structure that infants may initially recognize in the consonants and vowels of the ambient language. But how is it that the infant moves from the surface phonetics to the underlying phonology? And how might the infant's progress on this front be reflected in changing perceptual responses to nonnative phonetic patterns?

IV. ON ACCOUNTING FOR DEVELOPMENTAL CHANGES IN PERCEPTION OF PHONETIC INFORMATION

Two comprehensive, but radically different, theoretical approaches stand out in the scientific literature as providing possible accounts of how infants become attuned to the phonetic properties of their native language and begin to sort out the phonetics–phonology relations. The first approach is Noam Chomsky's linguistic theory of the grammatical structure of language and of its implications for language acquisition. Chomsky's premise of an innate Language Acquisition Device (LAD) is probably the most well known and widely accepted nativist perspective on language development. It is probably less widely known that his LAD was meant to apply to phonological as well as syntactic processes. The second is a psychological theory that is rarely applied to language or its development—James and Eleanor Gibsons' ecological perspective on perception. Their notion of perception as information pickup would suggest, as an alternative to an innate linguistic device, that perceptual learning may be the means by which language experience affects perception of native versus nonnative phonetic information.

To provide the foundation and rationale for the Perceptual Assimilation Model (PAM) of language-specific effects on speech perception to be presented in the subsequent section, this section critically examines Chomsky's and the Gibsons' theoretical approaches. It argues that whereas Chomsky's theory has provided important insights about the grammatical structure of language, including its phonological properties, some of his basic claims about the phonetics–phonology relation have not been supported by subsequent work in phonology. More important, difficulties with his nativist perspective on development lead me to reject that view as an approach to understanding the development of language-specific effects on speech perception in favor of the perceptual learning approach outlined by the Gibsons.

Following this theoretical discussion, PAM is developed as a perceptual learning account of listeners' perception of nonnative contrasts according to their phonetic similarities and dissimilarities vis-à-vis native phonological categories. The model is based on the principles of information pickup and perceptual learning put forth in the ecological theory of perception, as applied to listeners' recognition of language-specific relations between surface phonetic details and the underlying phonological principles that have been characterized by linguistic research. The model is discussed in light of recent cross-language perceptual findings with infants and adults from my own and others' laboratories. In addition, PAM's implications for the development of phonological knowledge about the native language is considered.

I now turn to an evaluation of Chomsky's proposal about language acquisition and of the Gibsons' theory of perception and perceptual learning. This discussion provides the groundwork for PAM.

A. Chomsky and the Language Acquisition Device

To set the stage, consider a quote from Chomsky's *Language and Mind* (1972), which illustrates his reasoning about the need for a language acquisition device. This particular passage was chosen because of its emphasis on the role of the LAD in phonological development:

> We can provide an explanation for a certain aspect of perception and articulation in terms of a very general abstract priniciple, namely the principle of cyclic application of rules. It is difficult to imagine how the language learner might derive this principle by "induction" from the data presented to him. In fact, many of the effects of this principle relate to perception and have little or no analogue in the physical signal itself, under normal conditions of language use, so that the phenomena on which the induction would have been based cannot be part of the experience of one who is not already making use of the principle. . . . Therefore, the conclusion seems warranted that the principle of cyclic application of phonological rules is an innate organizing principle of universal grammar that is used in determining the character of linguistic experience and in constructing a grammar that constitutes the acquired knowledge of language. (p. 45)

As indicated, a core premise of Chomsky's theory is that humans possess an innate biological specialization for learning language. This specialization is devoted solely to determining the specific grammatical structure of the native language, within the innately specified constraints on possible human grammars, on the basis of spoken input. The biological device—the LAD—is endowed with the universal grammar, that complement of grammatical functions found universally across languages. Thus, it includes the mechanisms that generate the language-specific rules by which the surface phonetic representations of utterances are derived from the underlying deep structure or abstract phrasal organization of intended meaning. Cross-language similarities in the structure of children's early grammatical constructions, their common phonological simplifications in pronouncing early words, and the disparity between those childish constructions and the grammars of the adult languages are taken as evidence for an innate biological specialization for language acquisition. The LAD makes possible the child's construction of a representation of the grammatical system of the native language, which includes the phonological rules by which sound and meaning are related, as can be seen in the following quote:

> The child constructs a grammar—that is, a theory of the language of which the well-formed sentences of the primary linguistic data constitute a small sample. . . . A child who is capable of learning language must have (i) a technique for representing input signals, (ii) a way of representing structural information about these signals, (iii) some initial delimitation of a class of possible hypotheses about language structure, (iv) a method for determining what each such hypothesis implies with respect to each sentence, (v) a method for selecting one of the (presumably, infinitely many) hypotheses that are allowed by (iii) and are compatible with the given primary linguistic data. (Chomsky, 1965, pp. 25–30)

Although his work on syntax is more extensive and more widely known outside of linguistics than his work on phonology, it is important to note that Chomsky considered the phonological patterning of a language to be a component of its grammar. Therefore, the endowment of the LAD also had to include the universal set of phonetic features—the full range of possible speech sound features from which all languages select a subset for the surface phonetic representation of utterances. The next quote, from *The Sound Pattern of English* (Chomsky & Halle, 1968; henceforth referred to as SPE), describes the predicted effects that knowledge of a particular language should have on the perception of phonetic features in speech:

> The hearer makes use of certain cues and certain expectations to determine the syntactic structure and semantic content of an utterance. . . . A person who knows the language should "hear" the predicted phonetic shapes. . . . Notice, however, that there is nothing to suggest that these phonetic representations also describe a physical or acoustic reality in any detail. . . . Accordingly,

> there seems no reason to suppose that [even] a well-trained phonetician could detect such contours with any reliability or precision in a language that he does not know. (pp. 24–25)

Thus, Chomsky and Halle posit that a listener's perception of phonetic patterns is determined by the phonological component of the specific grammar of his or her native language, *once the listener knows the language.* But if only a person who knows the language hears the phonetic shapes predicted by the grammar—the meaningful contrasts and phonetic equivalencies within its phonological component—then how should those same phonetic patterns be perceived by someone who does not know the language? More specifically, how does perception of the phonetic details of an unknown language differ between a listener who knows at least one language (i.e., knows a different language-specific grammar) and a listener who has not yet learned a first language (i.e., does not yet know a particular grammar)? How, indeed, does the first language learner acquire the native phonology, based on the spoken input from his or her language environment?

The answer to the last question, according to Chomsky, is that the LAD helps young children to determine the language-specific grammatical operations that relate the surface phonetic forms of native utterances to their underlying phonological, syntactic, and semantic representations. Because young infants innately possess the set of universal phonetic features, they should perceive the full range of possible surface phonetic contrasts in nonnative as well as in native speech. In this way, they remain open to learning whichever language is presented to them. But why, then, do not adults and older children also perceive the universal phonetic features in nonnative speech? The brief treatment of this issue in SPE points to the answer. It cannot be that mature language users have somehow lost the universal phonetic features with which they were born. Rather, it must be that, for them, the language-specific grammatical rules they have come to possess necessarily translate the surface phonetic features of utterances to the underlying phonological representations that are in accord with the grammatical principles of their language(s). That is, once the child has determined the rules of the language-specific grammar, he or she will "hear" the phonetic shapes predicted by the phonological component of that grammar.

This process would not constrain young infants' perceptions because they have not yet accrued sufficient language input to determine the underlying language-specific phonological representations of the ambient language's grammar. The LAD and its universal grammar are, nonetheless, present and operating even in the young infant. Its function in phonological development at this early stage is to construct the underlying grammar of the phonological component of the language by generating and testing hypotheses that could account for the observed patterning of the surface phonetic details in ambient speech.

To understand how this was expected to take place, we must briefly examine Chomsky and Halle's basic assumptions about how phonetic details relate to phonological representations. The classic view of SPE was that each consonant and vowel in an utterance is a discrete segment, represented phonologically as a feature matrix of all and only those phonetic features that distinguish it from all other segments in the language's inventory. The role of the phonological component of the grammar is to assign a language-appropriate phonetic feature matrix for the surface structure of each utterance generated by the syntactic component of the grammar. Thus, the phonological mapping to phonetic features is a part of the language-specific grammar. But the phonetic features are assumed to be binary, abstract, and timeless representations, even though their physical articulatory instantiations extend over time and space and show graded variability. That is, each static phonetic feature in a segmental matrix has only a positive (+) or a negative notation (−); the values for all features hold absolutely and concurrently in a segmental representation that has no time dimension. These static, binary feature specifications of the surface phonetic representation are automatically translated into the continuous, scalar articulatory details of real utterances, with temporal and spatial extent, by the universal grammar. That is, the translation to physical articulations is *not* part of the language-specific grammar. For these reasons, phonological representations do not incorporate all of the actual articulatory details associated with particular physical instantiations, such as the full range of details for specific dialectal or allophonic variants of a given segment. The latter sorts of detailed descriptions might be provided (by phoneticians) to fully characterize allophone-specific, dialect-specific, or even language-specific properties of utterances. But these would not be part of the language-specific grammar and so are not essential descriptions of phonological segments, which are abstract. Phonological segments represent the functional patterning of sound by the language's grammar and therefore are blind to allophonic or dialectal differences, which are phonologically equivalent in the underlying representation.

It is important to point out, however, that this segmental or linear view of phonology as propounded in SPE has largely been supplanted more recently by nonlinear or autosegmental phonology (e.g., Archangeli, 1988; Archangeli & Pulleyblank, in press; Clements, 1985; Keating, 1988, 1990; McCarthy, 1988; Prince & Smolensky, 1993; Sagey, 1986; for an introduction to autosegmental phonology, see Goldsmith, 1976). The nonlinear approach has developed in response to several difficulties with the classic linear model's handling of certain aspects of phonological patterns and phonetic implementations across languages. For one, the SPE claim that all features are binary fails to account for certain phonological processes; the nonlinear approach instead recognizes multivalent settings for certain phonological features. For another, the exclusively segmental domain of the SPE model failed to coherently incorporate certain effects of stress patterns, intonation, and syllable structure (phonotactics) on segmental properties. These effects are handled in nonlinear accounts by assuming instead that segments, stress, tonality, and syllable organization are distinct but interacting subcomponents of the phonology (e.g., Ito, 1986; Leben, 1978; McCarthy, 1986, 1987; Pierrehumbert & Beckman, 1988).

Another common phonological pattern is that phonetic features of one segment often "carry over" to other segments in an utterance, for example, vowel harmony or context-conditioned allophones. Because SPE-assumed phonetic features are linked to individual phonological segments, these phenomena required a proliferation of rules for moving phonetic features between segments. In nonlinear phonology, the effects follow automatically from an assumption that all features are independent of specific segments, with possible associations to one or more segmental "slots" (e.g., Cohn, 1990; Goldsmith, 1976; Inkelas & Leben, 1990; Kahn, 1980).

Finally, language- and dialect-specific differences in productions of segments with identical phonetic feature specifications call into question the SPE argument that articulatory implementation of phonological representations is automatic and universal, suggesting instead that articulatory details are part of language-specific grammar (see Fourakis & Port, 1986; Keating, 1988, 1990; Mohanan, 1986). For example, the ejective stop /p'/ is released later and hence more forcefully in Navajo than in Quechua (Lindau, 1982); nasal vowels have more delayed nasalization in Canadian French relative to continental French (van Reenen, 1982).

Although it has gone beyond the SPE model in handling certain phonetic and phonological patterns, however, the nonlinear approach has apparently retained the other basic theoretical premises of SPE. The nonlinear approach still assumes that phonological features are abstract and timeless. Moreover, nonlinear phonology proponents have had very little to say about ontogenetic development, certainly nothing that differs substantively from Chomsky's nativist assumptions (e.g., Archangeli & Pulleyblank, in press). That is, the nonlinear approaches retain, either tacitly or explicitly, the notion of an innate language acquisition device containing a universal grammar, with universal phonetic features.

However, those unquestioned assumptions, particularly certain assumptions underlying the posited innate linguistic device, raise some vexing problems. In-depth critiques of Chomsky's general theoretical framework have been offered from a linguistic perspective by Derwing (1973) and Sampson (1980) and from a psychological perspective by Bohannon, MacWhinney, and Snow (1990), among others (see special issue of *Developmental Psychobiology, 23*(7), 1990, for debate on both sides of the innateness issue). For the purposes of the present discussion, I focus on one of those problematic assumptions from Chomsky's claims about the LAD, exemplified in the following quote. The notion it conveys, that the input from the environment is inadequate in itself to directly specify the grammar of a language to a learner, characterizes a broader epistemological paradox of historical concern to epistemologists and perception theorists:

> The native speaker has acquired a grammar on the basis of very restricted and degenerate evidence; the grammar has empirical consequences that extend far beyond the evidence. At one level, the phenomena with which the grammar deals are explained by the rules of the grammar itself and the interaction of these rules. At a deeper level, these same phenomena are explained by the principles that determine the selection of the grammar on the basis of the restricted and degenerate evidence available to the person who has acquired knowledge of the language, who has constructed for himself this particular grammar. (Chomsky, 1972, p. 27)

Chomsky asserts in numerous places in his writings that the spoken input from the language environment provides inadequate information about the underlying grammar of the language for the child to apprehend that grammar directly. As the argument goes, each utterance of adult models offers the young child only an incomplete glimpse of the grammar of the language; some utterances are even ungrammatical. Moreover, caregivers generally fail to provide the sort of negative evidence that would unequivocally refute any incorrect hypotheses the child might entertain about the grammar of the language (e.g., Marcus et al., 1992). In short, the input is a sample of utterances that, individually, are incomplete (and consequently, sometimes ambiguous) reflections of the underlying grammatical system and that, collectively, presents but a tiny subset of the infinite grammatically acceptable sentences that a native speaker–hearer could automatically understand and produce.

Thus, the input utterances are taken to be informationally inadequate to specify the grammar completely and uniquely. Therefore, the reasoning proceeds, the child must innately possess a specialized device to construct a model of the grammar and test hypotheses against this input. Because this sort of data base has the potential to permit a large number of logically possible alternative descriptions of a grammar, innate constraints on the forms of permissible grammars are posited to be built into the LAD. Although these arguments have been developed primarily to account for acquisition of syntactic processes, it is presumed that phonology is subject to the same general principles as syntax. The surface phonetic input inadequately specifies the underlying phonological system, therefore phonological acquisition must depend on innate mechanisms. In the remainder of the current discussion, comments about the acquisition of grammar refer primarily to the phonological component (see Dent, 1990, regarding similar criticisms of nativist claims about semantic and syntactic development).

Here is the crux of the paradox: The grammar of a language, including its phonology, must be shared sufficiently well by the members of the language community for them to understand each other's utterances. Chomsky's argument is that the child cannot get the grammar directly from the inadequate evidence provided by adult utterances and so must use innate linguistic mechanisms to determine the grammar. But how can a shared grammar be developed in this way—individual mind by individual mind—based on inadequate input? How could such private grammars ever be verified, given the presumed inadequacy of the utterances[3] that are the only direct evidence that speaker-hearers can present to one another? How could those private grammars become mutually adjusted so that their users would be speaker-hearers of the same language?[4]

Chomsky's solution apparently is that the basis of this mutual adjustment is the innate endowment of linguistic concepts in the universal grammar that all humans share. Those innate concepts are employed to generate and test hypotheses about the grammar of a language against the primary linguistic data each child receives. However, as Chomsky acknowledged, a given set of primary linguistic data usually will support multiple solutions. To keep this problem from getting out of hand, he proposed that the number of potential solutions is limited by innate constraints on permissible grammatical forms. Nonetheless, multiple grammatical hypotheses are still to be expected; the language learner must select the "best" of the possible grammatical hypotheses generated to account for the observed data. Evaluation criteria for choosing the best among a set of possible solutions generally rely on concepts such as elegance or simplicity, which can be notoriously difficult to define and reach a consensus about (see Anderson, 1985; cf. Jeffreys & Berger, 1992). Again, the handling of this

---

[3]The written form is another type of direct evidence that speaker-listeners can present to one another, but it is subject to at least the same limitations as the spoken form. Presumably, the evidence it carries about the underlying grammar would also be considered inadequate. In any event, normal children learn to read and write only after they have learned to talk, so the written form would generally not offer an alternative basis for language learning (see also Liberman, 1992).

[4]In fact, the relation between the individual speaker-hearer's grammatical knowledge (linguistic competence), the same speaker-hearer's actual language behavior (linguistic performance), and the community's shared language is a complex issue. Although the matter cannot be explicated here, the reader wishing further information is referred to, for example, Chomsky (Chomsky & Halle, 1968, 1972), Newmeyer (1980), Sampson (1980), and de Saussure (1959).

problem is attributed to innate mechanisms—the requisite linguistic evaluation criteria are part of the LAD. But the difficulties of this line of explanation remain, compounded by the fact that the linguistic data set each individual receives will be different in particulars from that received by each other individual, even within the same community. Given this fact, how would the individual children of a language community end up generating and selecting the same, or similar enough[5], grammars?

All normal children, and many who are exceptional in some way, acquire the language spoken to them within a few short years. If the similarities among their disparate input sets are sufficient for children of a language community to select the same (or quite highly overlapping) "most elegant" solutions from among the various alternative grammars that each one privately generates, then surely this must mean that the input from adults provides robust and consistent rather than inadequate evidence about the grammar of the language. Indeed, if this be the case, why must the children construct their own private grammars at all? Why not learn the grammar directly from the patterning of the publicly available information in utterances, that is, learn the phonological system directly from the surface phonetic patterning of utterances?

The problem just summarized reduces to the philosophical paradox inherent in indirect theories of perception. The paradox has been recognized historically even by proponents of indirect theories. Specifically, it is that if inputs convey inadequate veridical information about the world, then we cannot directly know the outer world. The notion that we must know the world only indirectly, through deduction and interpretation of inadequate input, comes down to a claim that we can perceive in the world only what we already know is there to be perceived. This is, of course, the reasoning behind the standard nativist claim for innate knowledge. And as James Gibson (1979) argued, it is circular reasoning:

> Note that categories cannot become established until enough items have been classified but that items cannot be classified until categories have been established. It is this difficulty, for one, that compels some theorists to suppose that classification is a priori and that people and animals have innate or instinctive knowledge of the world. The error lies . . . in assuming that either innate ideas or acquired ideas must be applied to bare sensory inputs for perceiving to occur. . . . Knowledge of the world cannot be explained by supposing that knowledge of the world already exists. (pp. 252–253)

The claim for innate ideas would also seem to be at odds with the basic evolutionary principle of natural selection, dependent as that principle is on

[5]Indeed, how could one define "similar enough" if the utterances that serve as the only direct interface between different individuals' grammars inadequately reflect those grammars and thus are by definition inadequate to validate or reliably compare them?

the organism's fit to an ecological niche. That is, a species' survival is optimized when its physical structure and behaviors are well suited to those veridical properties of its world that are relevant to satisfying its procreative and survival needs. I argue that as applicable as these concerns are for indirect theories of perception of the physical world, they apply equally to Chomsky's nativist model for acquisition of the phonological grammar of a language. In particular, they are directly relevant to the assumptions that model makes about indirect perception of phonetic patterns in speech.

A fundamental problem of the indirect perception view is that it conceives of input to the perceiver from the world as a series of instantaneous collections of stimulus features that impinge on the special sensory organs (i.e., eyes, ears, nose) and that inadequately specify their dynamic and substantive sources in the world. Like snapshots, these inputs individually have no extension in time or space. A somewhat analogous view can be found in the nativist linguistic assumptions about the language input to the child, which could be characterized as *sound bites* of language—individual utterances each of which can provide only partial evidence about the underlying grammar, including its phonological component. According to indirect perception theories, because the stimulus cues are impoverished with respect to real-world events and objects, the perceiver presumably must use additional mechanisms of brain and/or mind to further process the sensory inputs, deduce what their sources must have been, draw inferences, develop memorial associations, and so on, in order to mentally construct an indirect representation of the world. But how could such mechanisms ever have evolved, given that the presumed inadequacy of the input would make it impossible for their outputs ever to be verified vis-à-vis the real world?

It was in response to these and other sorts of concerns about indirect theories of perception and perception-dependent knowledge in general that the Gibsons formulated an alternative, ecological approach to perception and perceptual learning (E. Gibson, 1969; J. Gibson, 1966, 1979). They argued that all animals, for the sake of their survival, must know the world directly from information available in stimulation.

B. The Direct Realism Alternative: Gibson's Ecological Theory of Perception

The ecological theory of perception represents the opposite philosophical extreme from the nativist assumptions of Chomsky's theory. The philosophical stance taken by the Gibsons' ecological theory of perception is that of direct realism, as opposed to indirect or innate knowledge. As the quote below illustrates, ecological theory assumes that stimulation is structured and dynamic, extending over time and space, and that it is directly detected rather than "interpreted" by innate knowledge, computation, inference, stored memories, or arbitrary associations:

> The evidence . . . shows that the available stimulation surrounding an organism has structure, both simultaneous and successive, and that this structure depends on sources in the outer environment. If the invariants of this structure can be registered by a perceptual system, the constants of neural input will correspond to the constants of stimulus energy, although the one will not copy the other. But then meaningful information can be said to exist inside the nervous system as well as outside. The brain is relieved of the necessity of constructing such information by any process—innate rational powers, (theoretical nativism), the storehouse of memory (empiricism), or form-fields (Gestalt theory). The brain can be treated as the highest of several centers of the nervous system governing the perceptual systems. Instead of postulating that the brain constructs information from the input of a sensory nerve, we can suppose that the centers of the nervous system, including the brain, resonate to information. (J. J. Gibson, 1966, p. 267)

As this passage indicates, information about the external world—about distal events, surfaces, and objects—is assumed to be directly picked up from stimulation by integrated perceptual systems. To illustrate the perceptual system concept, the retina of the eye does not gather visual information by working in isolation. Rather, it is an integral part of the perceptual system for seeing: two movable eyes fixed in a head that is attached to a body that can move to shift location and orientation of the viewer with respect to the external spatial layout; these components are neurally integrated with one another and with higher centers in the brain. Thus, the perceptual systems are assumed to have evolved to permit active, physical exploration of the world in the service of gathering and disambiguating distal information.

Thus, the ecological approach, like the linguistic nativist approach espoused by Chomsky, is concerned with biological specialization. However, the two views differ dramatically in their assumptions about the nature of biological specializations—the information they handle, the way they work, and the forces behind their evolution. According to ecological theory, the biologically specialized perceptual systems have evolved, and continue to function, for the pickup of veridical information from the world. This view admits the possibility of perceptual systems being specialized for pickup of information about specific types of distal objects or events, such as the information in speech that specifies the configuration and movements of the vocal tract producing the signal (see Best, 1984, 1993, in press-a, in press-b). Such specializations may be abstractly analogous to that of the human hands for grasping and manipulating objects and the complementary perceptual ability to detect the graspability and manipulability of distal objects. Evidence for primitive components of the latter abilities and of their responsiveness to the physical properties of distal objects (size, distance, speed of movement) is found quite early in development (e.g., von Hofsten, 1980). As for the pickup of distal articulatory information in the speech signal, Gibson summarized in general terms how and why this should be possible (see also Best, 1984, in press-a, in press-b; Fowler, 1986, 1989, 1991):

> An articulated utterance is a source of a vibratory field in the air. The source is biologically "physical" and the vibration is acoustically "physical." The vibration is a potential stimulus, becoming effective when a listener is within range of the vibratory field. The listener then *perceives* the articulation because the invariants of vibration correspond to those of articulation. In this theory of speech perception, the units and parts of speech are present both in the mouth of the speaker and in the air between the speaker and the listener. Phonemes are in the air. They can be considered physically real if the higher-order invariants of sound waves are admitted into the realm of physics. (J. J. Gibson, 1966, p. 94)

The direct realist philosophy assumes that information from the world is a rich multimodal flow of temporally and spatially distributed energy patterns that are lawfully and systematically shaped by distal events and objects. The systematic structure in this information flow is picked up by perceptual systems—extracted, detected, discovered—through active, physical exploration of the events, surfaces, and objects that shape the energy flow. By shifting position and orientation with respect to the objects and the spatial layout, as well as by moving and manipulating objects, the perceiver produces changes in the flow of stimulation that are systematically influenced by the exploratory actions in ways that provide rich, direct, veridical information about the distal sources of stimulation. As a result of this active exploratory behavior of the perceptual systems, the perceiver becomes better attuned, with increases in experience, to the invariants in stimulation that specify the defining characteristics of specific events, the persisting identity of particular objects, and the higher order commonalities shared by similar events or by similar objects.

The transformational invariants of an event are those properties of the energy flow that remain constant across the participation of different objects in that event. For example, the transformational invariant of repetitive rotation about an axis specifies the same event of spinning, whether a top is spinning on a surface, an amusement park "antigravity" ride is spinning to produce centrifugal force, or the wheels of a car are rotating on their axles. The structural invariants of spherical shape and elastically deformable solid specify an identity relation—the same baseball across the events of rolling, throwing, bouncing, and juggling. Invariants can also specify similarity relations among objects or events. The more abstract invariant of a convexly curved plane characterizes the primary similarity among the outer surface of an eyeglass lens, the dome of an enclosed sports arena, and the silhouette of an old Volkswagen "beetle." And although the following do not reflect literally the same event, they involve abstractly similar curvilinear movement transformations: the slithery, winding progression of a snake, the sinewy movements of a traditional Thai dance, and the wavelike motion of tall grass rippling in a breeze (for further discussion of structural and transformational invariants, see Shaw, McIntyre, & Mace, 1974). Experience-dependent changes in attunement to such invariants occur through perceptual learning.

The ecological perspective has concerned itself primarily with general perceptual principles rather than with linguistically specialized mechanisms. However, I believe it is eminently applicable to children's learning of the sound pattern of their native language and to the concomitant effect of this learning on the perception of nonnative sounds and contrasts. If we take an ecological view of the realm of language, the spoken input available to the young child is a flow of many utterances, occurring multimodally within a rich behavioral context that extends over time and people. The flow of this linguistic and social stimulation, extending as it does over time and speakers, should reveal regularities or invariants across utterances that the infant comes to recognize as the sound-organizing principles of the phonology of the language (e.g., Best, in press-a).

I have taken the ecological perspective to account for how experience with the ambient language comes to influence the infant's perception of nonnative speech contrasts. To do so, I apply this perspective to linguistic insights about the sound structure of languages, which should form the basis for the child's developing recognition of the relations between the phonetic properties of speech and the phonological organization of the grammar of his or her native language. For the purposes of this chapter, I am particularly interested in how ecological principles apply to perceptual learning, specifically with respect to infants' and young children's perception of the sound pattern of their native language. Therefore, I now turn to examine in greater depth the ecological approach to perceptual learning.

C. The Ecological Perspective on Perceptual Learning

Two quotes exemplify the ecological viewpoint on perceptual learning: the first is from James Gibson's (1979) book, *The Ecological Approach to*

Visual Perception, and the second is from Eleanor Gibson's address, "Perceptual Development and the Reduction of Uncertainty" at the 18th International Congress on Psychology in Moscow:

> The perceiving of the world begins with the pickup of invariants. . . . [T]he theory of information pickup . . . needs to explain learning, that is, the improvement of perceiving with practice and education of attention. . . . The state of a perceptual system is altered when it is attuned to information of a certain sort. The system has become sensitized. Differences are noticed that were previously not noticed. Features become distinctive that were formerly vague. (J. J. Gibson, 1979, p. 254)

> Discrimination learning proceeds . . .by discovering distinctive features of objects and invariants of events in stimulation. . . . The effective stimulus which active and educated perception picks out is a reduced stimulus. It is extracted, filtered out, whereas other stimulus information which has no utility for differentiation is ignored by the educated attention. (E. J. Gibson, 1966, pp. 10-15)

When a perceptual system becomes attuned to a particular type of information, it becomes altered by experience. The claim is that the attuned perceiver is more quickly and efficiently able to pick up from the flow of stimulation just that information to which the perceptual system has become sensitized, as opposed to, perhaps, simply increasing the speed of a cognitive search through mental space. This sensitization of the perceptual system entails detection of critical distinctions among objects or events that had previously gone unnoticed. What is suggested by perceptual learning, then, is an optimization and economization of pickup or extraction of critically distinctive properties. Perceptual learning is probably more readily apparent for detecting abstract, higher order invariants (such as the curvilinear movement invariant described earlier) than for detecting the simple, lower order invariants to which perceptual systems are innately tuned even very early in life (e.g., basic color categories; Bornstein, 1979).

These principles have been more completely drawn out by Eleanor Gibson in her numerous writings on perceptual learning (e.g., E. Gibson, 1963, 1966, 1969, 1977, 1988; E. Gibson & J. Gibson, 1972; J. Gibson & E. Gibson, 1955). As her opening quote indicates, perceptual learning leads to improved discrimination, but this does not mean simply the discrimination of smaller and finer stimulus differences, hence, of always increasing numbers of individual stimuli. Instead, perceptual learning entails the discovery, for specific purposes, of the critically distinctive features of objects and invariants of events in stimulation. It involves the education of attention for most efficient detection of the most telling differences among objects and events that are of importance to the perceiver. As she has argued, the utility that critical distinguishing features and invariants of events have for the perceptual learner is that they reduce uncertainty among choices in a world that otherwise presents too much, rather than too little, information. Educated attention, that is, a perceptual system that is attuned to certain types of information, picks up *reduced* stimulus information, which is selected, extracted, or filtered out from the larger flow specifically because of its ability to critically differentiate things that are of interest or usefulness to the perceiver. Other stimulus information that does not serve this purpose of utility is ignored, that is, not picked up.

This account leaves open the possibility for reeducation of perception, because the undetected information is still available in stimulation. Stimulus information that is irrelevant for well-used distinctions, and therefore has been systematically ignored, could later prove important for other new distinctions. It is conceivable, perhaps even likely, that having first learned to economize information pickup by overlooking certain information as irrelevant (or by perceiving it as equivalent to some other pattern of information) may make it more difficult to relearn to attend to it later than would be the case for a novice learning to attend to the same information for the first time. Ecological theory has not directly addressed these possibilities. However, they are relevant for understanding whether and to what extent second language learners may learn to detect nonnative phonetic distinctions that are not utilized in their native language and in what way this may be affected by varying degress of experience with the native language.

Indeed, the Gibsons did not address speech perception in great detail in their primary accounts of the ecological apoproach to perceptual learning (cf. E. Gibson & J. Gibson, 1972), although Eleanor Gibson did address certain aspects of language in her research on reading development (e.g., E. Gibson, 1971). The ecological view on perceptual learning has primarily addressed the general issues of how perception is shaped by experience. Perceptual learning entails the discovery of invariants in stimulation that reveal the structural and functional properties of the source objects and events. Often these invariants are hierarchically nested in complex events, so that higher order invariants may depend on, or be derivatives of, lower order invariants. Discovery of certain higher order invariants may thus be possible only once the perceiver has learned which of the lower order invariants are critical to the distinction and which are not. Perhaps, for some distinctions, there may even be several levels of lower invariants supporting the discovery of a higher order invariant.

Spoken language provides an excellent example of the sort of complex organization in which higher order invariants, such as those that specify syntactic principles, may not be detectable until the perceiver has learned to pick up certain distinctive information at lower levels, such as the critical differences in the phonetic patterns of similar sounding but meaningfully different words. For the infant, then, learning the sound pattern of the native language is the quintessential task of perceptual learning, that is, discovering the multiple levels of invariant principles by which the stimulus flow is patterned.

The ecological premise is that the complex, nested hierarchy of linguistic organization, including phonological patterning, exists in the infant's language environment. It is all there, that is, if we consider the available language stimulation to span the history of utterances the infant hears, along with the rich behavioral contexts in which those utterances occur. The flow of spoken utterances in context provides the infant a window on the patterning of the ambient language. This is the flow of stimulation from which infants must learn to recognize and abstract the invariants that specify all levels of linguistic structure. Of course, the infant is not initially able to detect or abstract from that flow the invariant properties specifying most of the levels of linguistic organization summarized earlier. In fact, the only level of available information that the infant is likely to be able to detect initially is the surface phonetic information. And it is necessary from among those phonetic details that the infant must learn to recognize the higher order invariant patterns that specify words, syntax, morphology, and, in particular, phonology.

Thus, the ecological view is that utterances provide a rich flow of information about dynamic speech events that extend over time, and that through perceptual learning the individual becomes attuned to various levels of invariant structure available in that flow. This view suggests a radical departure from the standard assumption of discrete, timeless features and calls instead for a model of phonetics and phonology in which the crucial dynamic attributes of events in the speech world are integral to the model. The ecological perspective has begun to offer alternative insights and evidence both about the phonetic details of speech production (Fowler, Rubin, Remez, & Turvey, 1980; Kelso, Saltzman, & Tuller, 1986; Saltzman & Munhall, 1989) as well as about its phonological organization (Browman & Goldstein, 1986, 1989, 1990a, 1990b, 1990c, 1992a; Fowler, 1980; Goldstein & Browman, 1986). The latter work has offered an articulatory gestural model of phonology, which I examine next as the basis for an ecological, perceptual learning account of language-specific effects on the perception of nonnative phonetic contrasts. The following summary is based on the works of Browman and Goldstein cited earlier.

V. GESTURAL PHONOLOGY

The tenets of gestural phonology are grounded in the spatiotemporal organization of articulatory gestures in speech, which are themselves grounded in the biomechanical organization of the human vocal tract. Rather than assuming abstract and timeless phonetic features as the atoms or primitives from which phonological representations are built, the gestural model assumes that the phonological primitives are articulatory gestures—the coordinated actions of vocal tract articulators. The model organizes these gestural features within the framework of a hierarchical *articulatory geometry* based on the anatomical relations among the articulators involved in speech. The vocal tract is comprised of three relatively independent articulatory systems that are represented as separate nodes within the articulatory geometry: the glottal system (vocal cords), the nasal system (the velum, the valve that permits or prohibits air flow through the nasal cavity), and the oral system (which includes the lips and the tongue as separate subsystems). There is an additional subordinate level in the tongue subsystem: tongue tip versus tongue body, whose actions are differentiated by different intrinsic and extrinsic muscles of the tongue. This hierarchically organized set of artculators functions within the confines of the walls of the vocal tract, which is structured basically as a bent tube of varying diameter, optionally connected to a second side tube (nasal cavity) via the open velum. The coordinated actions of the articulators can cause constrictions at various locations (place of articulation) along the vocal tract (e.g., dental, alveolar, velar, etc.; see Figure 1 for additional places of articulation). Each place can display several variations in degree of constriction, which determines the manner of the sound produced (complete closure for stop consonants, critical constriction for causing turbulent airflow in fricatives, narrow constriction for some vowels and for approximant consonants such as /w/ and /r/, wide opening for the velum in nasals and the glottis in voiceless sounds). Articulatory geometry is compatible, in many respects, with the nonlinear or autosegmental approaches that have supplanted SPE phonology. Some important distinctions must be noted, however, between the two approaches. Specifically, gestural phonology posits phonological elements to be gestures defined by a set of dynamic equations describing the movement of articulators over space and time, rather than a specification of abstract, timeless phonetic features. To illustrate, the equation set for the syllable *ma* describes a velum opening gesture and lip closing gesture that begin simultaneously and reach their peaks synchronously to produce the /m/ and a slower, less extreme tongue body gesture to narrow the pharynx (upper throat) for the "ah" vowel, which begins synchronously with the other two gestures but peaks later and lasts longer.

Thus, articulatory geometry is closely related to the anatomical structures and movement patterns of the vocal tract. This way, in the gestural model the phonological primitives and their physical instantiations derive from a single domain grounded in the spatiotemporal properties of real

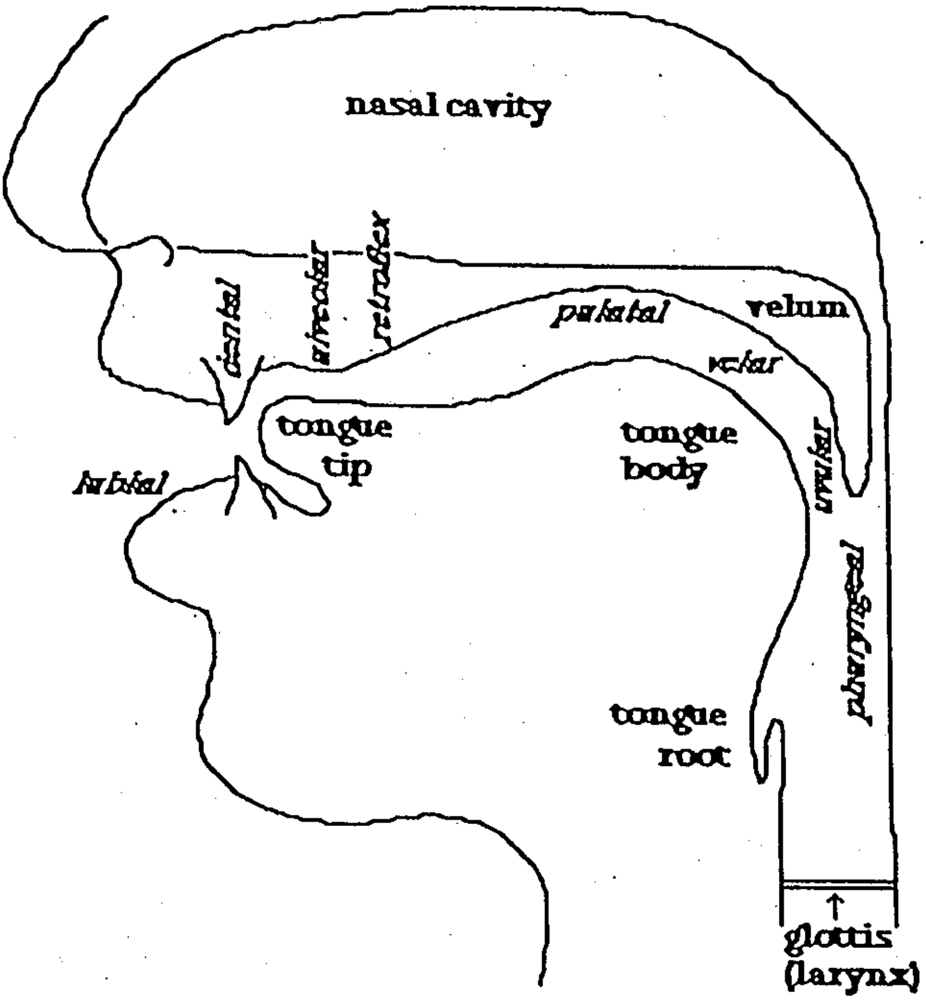

Fig. 1. Schematic lateral view of vocal tract, with major articulators labeled and the nasal cavity identified. Many of the common places of articulation, or locations of articulatory constrictions, are indicated in italics.

articulatory events. Because of this, phonological representations can specify the relative timing, or phasing, of one articulatory gesture relative to another. For example, the Canadian French versus continental French difference in vowel nasalization that was mentioned earlier (van Reenin, 1982) can be specified dynamically as a difference in the relative timing, or *phasing*, between the onset of velum lowering for nazalization and the peak of tongue movement for the vowel. This characterization departs critically from the phonetics–phonology relationship held by classic SPE phonology and by nonlinear phonologies, neither of which can phonologically represent the dialectal difference phonologically, even though the nasalization difference appears to be part of the language-specific grammar in the two dialects. This representational inability occurs for the latter two views because they posit that phonetic and phonological information exist in two divergent, informationally incompatible domains—one physical (actual articulations) and the other, only mental (underlying phonological representations).

In gestural phonology, the dynamical specifications of articulatory gestures describe change over time in particular vocal tract variables and their associated articulators (e.g., location and degree of a constriction by the tongue tip or tongue body somewhere along the vocal tract tube; opening of the nasal tube by movement of the velum). The model assumes that articulator motion is governed by dynamic principles of spring-like physical systems,[6] in which the values of several parameters of the tract variable(s) are specified: mass, stiffness, damping, rest position, instantaneous position, acceleration, and velocity. All tract variables are assumed to have a resting, or default, setting. The resting state is not, of course, specified as a gesture; gestures are active articulatory movements away from the resting state. A given gesture is a particular transformation of a tract variable (e.g., complete closure of the lips) that remains invariant across different contexts, speaking rates and styles, and speakers. There may also be variation in the exact articulators or coordinations among articulators that are used to achieve essentially identical gestural goals. For example, bilabial closure may be achieved by moving only the lips and keeping the jaw angle constant or by keeping the lips immobile and changing only the jaw position to bring the lips closer together (see Abbs & Gracco, 1984). Therefore, the dynamical description of a particular gesture defines a family of articulatory trajectories that all achieve the same gestural target of a particular degree of constriction at a particular location along the vocal tract tube.

Some phonological elements are composed of only a single gesture, whereas others involve a specific pattern of coordination between two or more individual gestures. Coordinations among two or more gestures are called *gestural constellations*. I illustrate the difference with the /p/–/b/ contrast, which in classic phonological description shares the phonetic features [+anterior], [−continuant], and [−sonorant] and are distinguished only on the feature [+/− voice]. But in gestural description, the voiced stop /b/ in *gabbing* involves only a single bilabial closure gesture (complete closure and release of constriction at the lips). The state of the *glottis*, or opening between the vocal folds, is maintained in the default adducted position (critical constriction rather than tightly closed) and produces voicing throughout the word. In other words, there is no active

---

[6]Currently, the model assumes that articulator movement is modeled fairly well by the dynamic regime of a *point attractor*, or damped mass spring, model with constant mass for each articulator. Such dynamic regimes characterize the pattern of movement of a physical system moving smoothly toward a single target (*attractor*).

glottal gesture just for the /b/. In contrast, the cognate voiceless stop /p/ in *gapping* involves two gestures that must be correctly phased relative to each other. Specifically, the bilabial closure must cooccur with an active glottal opening gesture, which prevents voicing and instead permits turbulent airflow (i.e., aspiration noise) through the vocal folds. The peak opening of the glottis coincides with release of the bilabial constriction; the glottis returns to its default state (vocal folds together for voicing) after bilabial release. The /p/ example illustrates a gestural constellation that corresponds to the segmental level of traditional phonology. But gestural constellations may also describe articulatory coordination at the level of syllables, words, prosodic phrases, and so on. Analogous to nonlinear phonological approaches, these nonsegmental levels of linguistic organization among gestures are specified for different articulatory tiers, such as those representing syllable structure and stress units. However, neither gestures nor constellations bear a one-to-one relationship either to segments or to classic phonetic features.

Because gestures are defined by a dynamical pattern of articulatory movements, each gesture has both an intrinsic spatial aspect and an intrinsic temporal aspect. This grounding in the physical properties of events over time departs qualitatively from the classic and the nonlinear views of static, dimensionless phonetic features. In gestural phonology, the phasing principles among the gestures in a given utterance are represented in both their spatial and temporal relations in a *gestural score*. To illustrate, a schematic gestural score for the word *mob* ([mɑb]) is shown in Figure 2. The abscissa represents the time line of the utterance; the ordinate represents the tiers in the articulatory geometry that are needed to display the critical gestures involved in that particular word. The rectangular boxes represent the temporal extent during which given gestures are active for their corresponding articulatory tiers or articulatory sets (e.g., tongue tip, tongue body, etc.). Inside each activation interval box, the degree of constriction achieved in the gesture and its specific location along the vocal tract are denoted. An American English utterance of *mob* begins as was described earlier for the syllable *ma*. The pharyngeal gesture for the vowel ("ah") extends into the final bilabial closure that corresponds to /b/.

Thus far, the gestural phonology approach has been applied in detail primarily to American English alone, but it can be extended (and in some cases has been) to suggest gestural characterizations of certain similarities and differences between the gestural constellations for some nonnative phonetic contrasts and contrasts found in the English phonological system. A few cross-language comparisons are offered here as illustrations. However, bear in mind an important caveat from Browman and Goldstein (1992b) that any proposed gestural analysis is obviously incomplete and speculative in the absence of hard data on the actual gestural processes

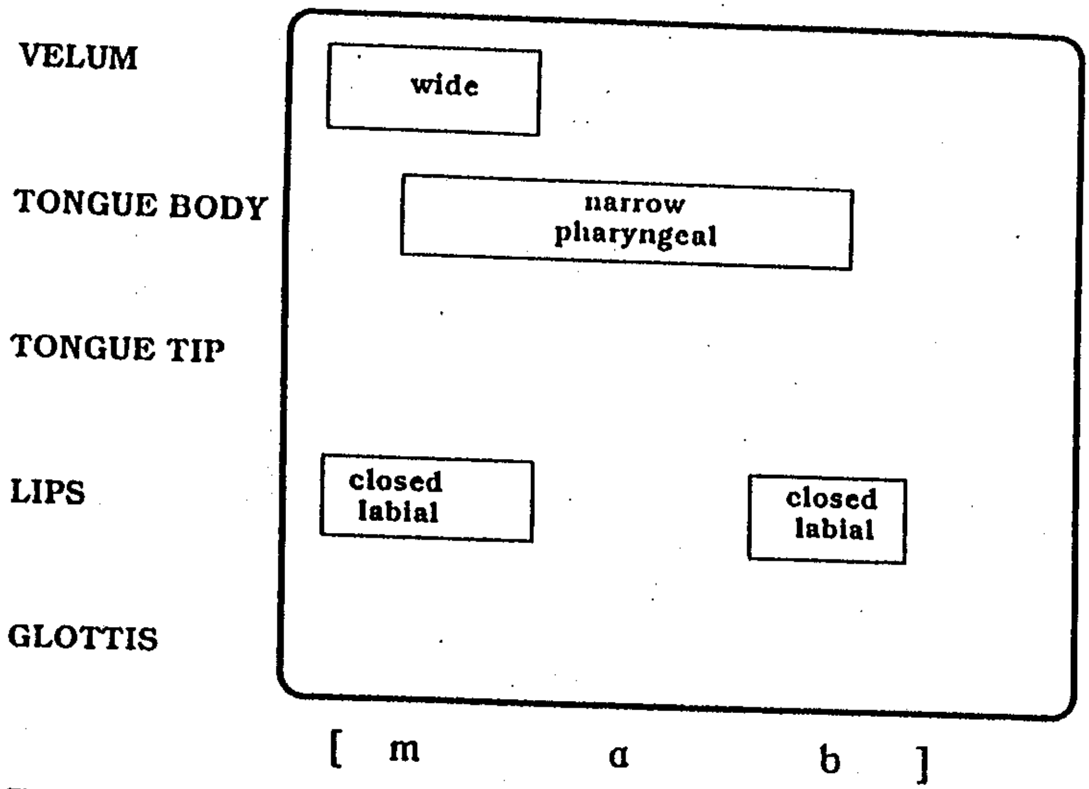

Fig. 2. Schematic gestural score for the word <mob> [mɑb] using box notation to indicate activation intervals for gestures and phasing among gestures.

involved in the utterances being considered. The comparisons here are based on currently available phonetic, acoustic, and physiological descriptions for the phonological contrasts involved. But the schematic gestural scores offered are necessarily speculative because of the incompleteness of actual gestural evidence, especially with respect to temporal extent and precise phasing of gestures.

Figure 3 shows the Hindi dental-retroflex contrast [d̪ɑ]-[ɖɑ] and English [d ɑ], which is gesturally most similar to both Hindi patterns. The schematized Hindi gestural scores and the English one are essentially the same except that the Hindi constriction locations are just anterior and just posterior, respectively, to the English alveolar location. Recall also that English does have context-conditioned dental and retroflex allophones of /d/, but not in the context of an isolated [dɑ]. Schematic gestural scores for the Zulu aspirated versus ejective velar stops [k^ha]-[k'a] are compared to the correspondingly most similar English gestural constellation: [k^ha], in Figure 4. In this case, the Zulu aspirated token is virtually identical to the English one, whereas the ejective token deviates from it in the constriction degree of the glottal gesture, which is closed rather than wide, producing silence rather than aspiration prior to the onset of voicing for the vowel. A different type of Zulu contrast is between voiced and voiceless lateral fricatives. These gestural constellations are produced with essentially the

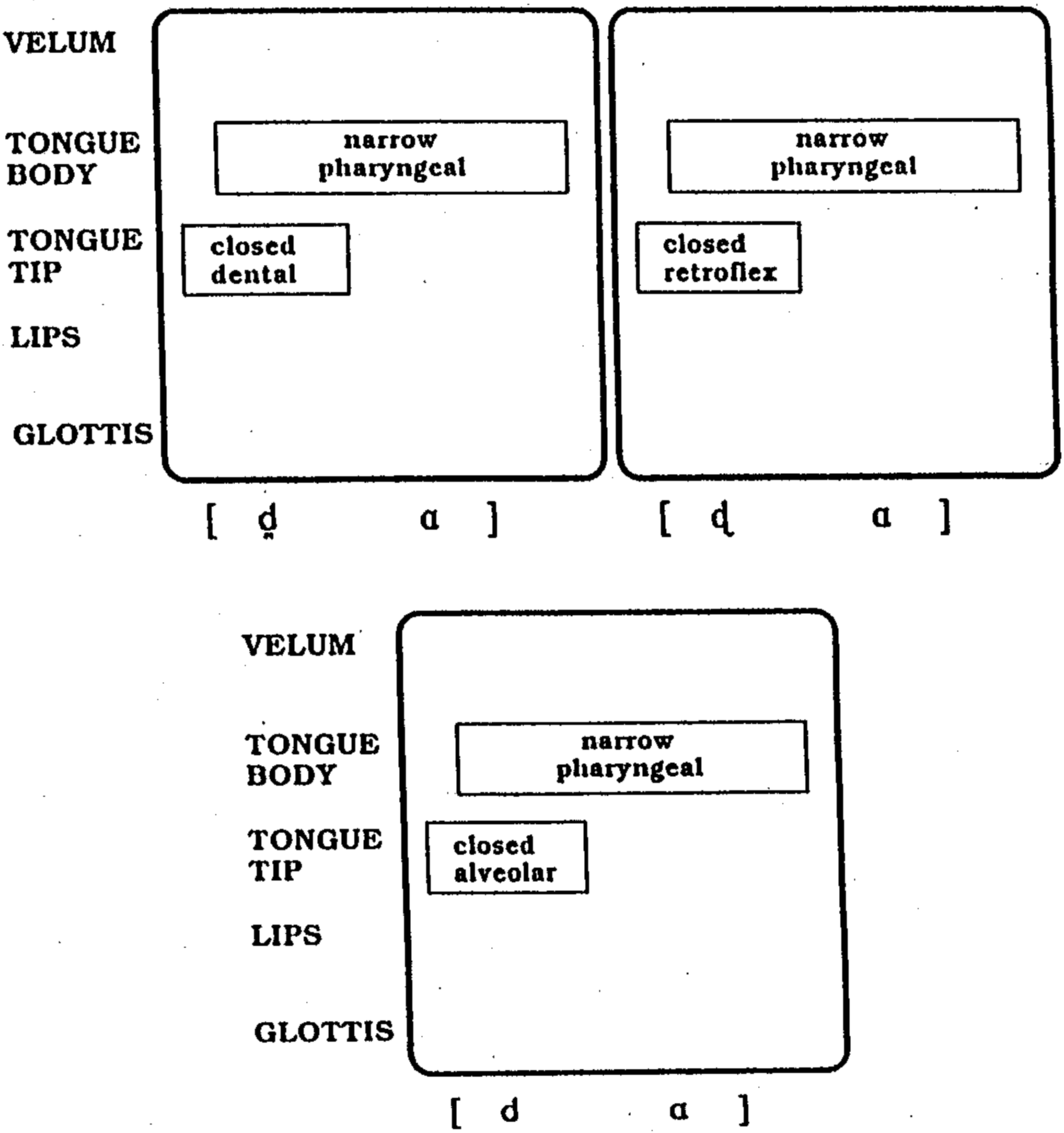

Fig. 3. Schematic gestural scores for the Hindi dental-retroflex /d̪ɑ/-/ɖɑ/ contrast (top panels) and English /dɑ/ (bottom panel).

same alveolar tongue tip closure and uvular tongue body narrowing as in English /l/. They differ, however, in employing a smaller constriction degree along the two sides of the tongue (against the upper lateral teeth) than for /l/. Instead, the lateral constriction is critical, producing airflow turbulence analogous to that at the tongue tip for fricatives such as English /z/ or "zh" or voiced "th" (in *that*) versus /s/ or "sh" or voiceless "th" (in *think*). Thus, the Zulu lateral fricatives gesturally resemble both the liquid /l/ and the voiced–voiceless fricative distinctions of English that involve tongue tip constrictions at anterior locations. Larger English gestural constellations (multisegmental) that may approximate the patterns found in the lateral

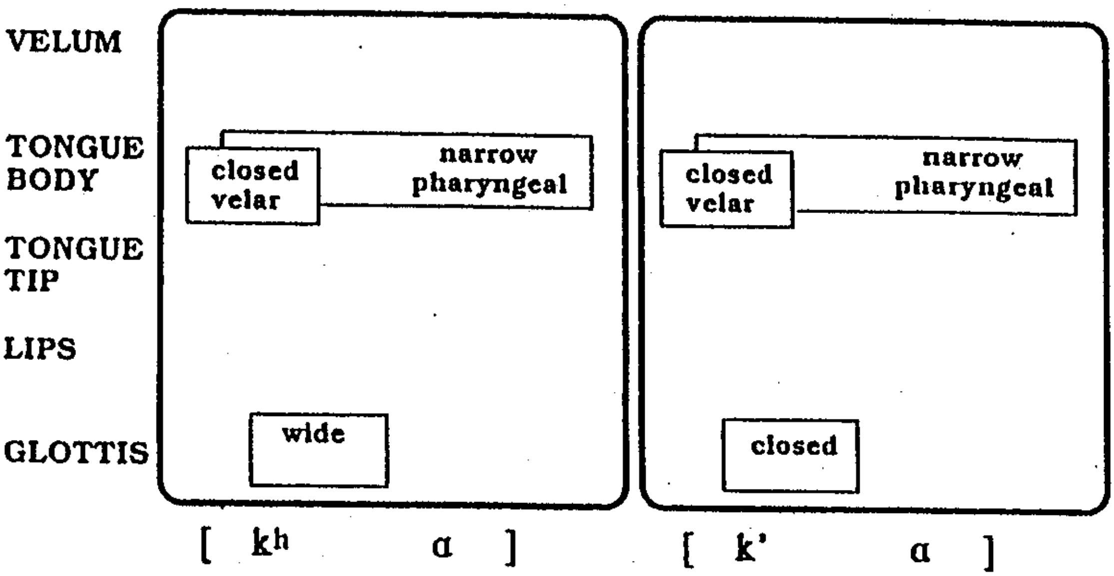

Fig. 4. Schematic gestural scores for the English and Zulu voiceless velar stop /kʰɑ/ (left) and for the Zulu ejective velar stop /k'ɑ/ (right).

fricatives include /z/–/sl/ *(paisley, slow)*, "zhl"–"shl" *(beigely, Ashley)*, or voiced versus voiceless "thl" *(blithely, breathless)*. Finally, the Zulu alveolar versus lateral click consonants incorporate gestural constellations that are quite dissimilar from any in English. Both have full closures at two locations: alveolar (tongue tip) and velar (tongue body). A vacuum is created in the intervening zone by drawing the tip or one side of the tongue downward until the suction is released. In syllabic context, this is followed immediately by release of the velar closure. The double closure plus suction release does not closely resemble any English gestural constellation.

Gestural phonology can also account parsimoniously for a wide variety of phonological phenomena within its articulatory framework, using gestural primitives that have intrinsic temporal and spatial dimensions, unlike static, dimensionless phonetic features. In most cases, these gestural accounts are backed by speech production data. For example, minimal contrasts are two gestural constellations that are identical except for a critical difference in constriction location (e.g., /b/ vs. /d/) or constriction degree (e.g., /b/ vs. /w/) in the oral tier of the articulatory geometry, presence or absence of a gesture of the velum (e.g. /ma/ vs. /b/) or glottis (/p/ vs. /b/), and so on. The tube geometry of the vocal tract also appears to account straightforwardly for certain natural classes, that is, groupings of different types of phonetic categories that nonetheless participate together in widespread phonological processes. To illustrate, nasals, liquids (/r/, /l/), and vowels form the class defined traditionally by the [+sonorant] feature, which has been difficult to define objectively. In gestural phonology, these phonetic types share the simple gestural similarity in which they all maintain one of the two vocal tract pathways (oral, nasal) wide open for outward airflow (Browman & Goldstein, 1989). Many allophonic variants can be explained as the overlapping of adjacent gestures, or *coarticulation,* as in the dental allophone for /n/ in te*n* themes, which results from overlapping the wide velum for /n/ and the dental location of the tongue tip for "th" (Browman & Goldstein, 1989). Analogously, gestural overlap can account for certain cases of phonological assimilation, as when the /n/ in seve*n* plus assimilates to /m/ in casual speech. The feature-based rule is that the labial feature of the /p/ spreads forward to the /n/. The gestural explanation is that the bilabial closure gesture of the /p/ overlaps the velum opening gesture for the /n/, thus "hiding" the aerodynamic evidence of the alveolar tongue gesture for /n/ and producing the bilabial nasal /m/ (Browman & Goldstein, 1989).

Cases of phonological deletion can be handled likewise from a gestural perspective. For example, feature-based approaches posit a deletion rule, whereby the final /t/ of the first word in *perfect memory* gets deleted, but in gestural terms it is simply the case that the alveolar /t/ gesture gets hidden by an overlap with the /m/ of *memory* (Browman & Goldstein, 1989, 1990c). Gestural overlap can even account for the insertion of an additional segment between other segments, called *epenthesis.* As an illustration, *something* is often pronounced in American English with a /p/ between the /m/ and the "th," leading feature-based accounts to invoke an insertion rule. But the /p/ arises gesturally from the overlap of the bilabial closure gesture for the /m/ and the glottal opening gesture for the following "th" (Browman & Goldstein, 1990c). The phenomenon of *metathesis,* in which the sequential order of segments becomes reversed by some phonological process, has been particularly vexing for generating feature-based rules that are powerful enough to describe the phenomenon, but not so overly powerful as to generate many nonoccurring reversals. Such ordering reversals often occur in speech errors, as when the rapid production of *Bob flew by Bligh Bay* comes out as *Blob foo by Bligh Bay.* A gestural analysis of tongue movement for the /l/s in these utterances reveals evidence of the temporal "sliding" or overlap of the tongue tip constriction gesture with those preceding it in the represented sequence, causing both overt and covert speech errors (Browman & Goldstein, 1992a).

The gestural phonology model has received some criticism from nonlinear phonologists, as well as some praise. On the positive side, some phonologists acknowledge that placing articulatory constraints on phonological processes is advantageous (see also Archangeli, 1988; Archangeli & Pulleyblank, in press), especially with respect to better delineation of the relation between phonology and phonetics (e.g., Clements, 1992; Pierrehumbert & Pierrehumbert, 1990). By and large, the criticisms reflect two underlying observations: (a) gestural phonology rejects static, timeless phonological features that differ in kind from physical, phonetic realizations; and (b) it does not invoke abstract cognitive rules about phonological representations (e.g., Pierrehumbert, 1990; Pierrehumbert & Pierrehumbert, 1990; Steriade, 1990). In other words, gestural phonology rejects two central tenets held by both SPE and nonlinear phonologies. These criticisms also suggest some partial misunderstanding of gestural phonology. The model does include discrete, or categorical, elements at the phonological level of the task dynamics used to generate gestures (Browman & Goldstein, 1992b). Moreover, it does distinguish between phonological and phonetic levels of representation, but views them as macroscopic versus microscopic descriptions of the same dynamic, physical domain of speech events (Browman & Goldstein, 1990a; see also Ohala, 1990). This brings us back to the central claims of the ecological approach, which assumes that perception must be grounded in physical reality. On that note, I return to the issue of how the physical properties of native speech are perceived by the adult and learned by the child.

VI. THE ECOLOGICAL APPROACH TO PERCEPTUAL LEARNING OF SPEECH

All of the phonological approaches discussed, including gestural phonology, have taken their task to be the generation of a physical phonetic output from the more abstract phonological component of the grammar. But I began with, and now return to, the opposite process—how a perceiver, particularly a young learner, gets from the phonetic surface to the phonological structure via perception. Specifically, the chapter began with the question of how experience with one's native language comes to affect one's perception of nonnative speech sounds and contrasts from unfamiliar languages. Phonology has provided little guidance here. Although Chomsky and Halle stated in SPE that a listener who knows the language being spoken will hear the phonetic shapes predicted by the phonology, it is unclear how they would expect the phonology to handle discrepancies between the phonetic features in a nonnative sound and the feature matrices defined by the phonological system of the listener's language. Indeed, how would it even handle perception of corrupted native speech (e.g., foreign accented or disordered speech) or the phonetic patterns of an unfamiliar dialect? Nonlinear phonological approaches do not help much, as they also have devoted minimal attention to theoretical issues in perception. And gestural phonology, the youngest of the approaches, has also focused the majority of effort on production. Moreover, none of these phonological approaches has given any depth of consideration to how infants and young children perceptually learn about the phonological structure of their native language.

To address these issues, I return to the direct realist view of speech perception based on the Gibsons' ecological theory of perception. This view assumes that listeners perceive information in speech about the distal articulatory gestures that shaped the phonetic patterns (Best, 1984, 1993, in press-a, in press-b; Fowler, 1986, 1989, 1991). Because it assumes that phonological processes derive from the same physical, dynamic domain as the phonetic details of actual utterances, gestural phonology lends itself to an ecological perspective on cross-language influences in perception as well as on how the infant learns the phonological properties of the native language. Articulatory gestures would provide a common metric for both perception and production of speech. The interrelation of perception and production is central to both speech imitation and language acquisition.

The direct realist view posits that perceivers recover information from speech and from other sound-producing events about the distal structures and events that produced the sounds. This view assumes that information about articulatory gestures is directly perceived in speech, as opposed to being the end product of cognitive processing of the raw acoustic input. The speech signal is shaped by the structure and movements of the vocal tract according to physical laws, as indicated by the earlier quote from James Gibson. Thus, evidence about articulatory gestures is available to perceivers as structured information about the speech events that produced the signal. This view is *not* the same as that of the well-known motor theory of speech perception (e.g., Liberman & Mattingly, 1985), which posits that perceivers refer to the motor control of their own speech in order to perceive the phonetic structure of speech input. The ecological claim is that listeners perceive the speaker's articlatory gestures as such, without referring to their own articulatory commands and, indeed, regardless of whether they can themselves produce similar signals.

That listeners perceive gestural information in speech is supported by cross-modal speech perception research (see also Best, 1993; Studdert-Kennedy, 1993). McGurk (McGurk & MacDonald, 1976) found that when presented with audiovisual syllables in which the synchronized consonants in the two modalities are from different categories, listeners perceive a unified phonetic pattern that is compatible with both modalities, rather than noticing the discrepancy. That is, the two modalities apparently provide evidence about a common, underlying dimension such as articulatory gestural patterns. An alternative argument—that the perceptual link between visual and auditory information is learned by association—is illogical in the general case, according to the Gibsons' arguments, and has been empirically refuted for the speech perception case by two recent reports. Cross-modal integration does occur for synchronized but discrepant consonants presented auditorily and tactually—blindfolded subjects manually felt the movements of an experimenter's silent lip movements synchronized with audio recordings, although they had never had such tactile-auditory experience with speech. Yet, there was no cross-modal integration for synchronized audio and written syllables in the face of the subjects' extensive associative experience with the relation between text and speech (Fowler & Dekle, 1991). In another study, young English-learning infants heard repetitive audio presentations of the French lip-rounded vowel /y/, which does not occur in English, synchronously with side-by-side silent videos of the English lip-rounded long "oo" and unrounded "ee" (Walton & Bower, 1993). The infants preferentially fixated on the "oo" video when hearing /y/. Given their lack of prior experience with /y/, this could not have been a learned association but, rather, suggests detection of the articulatory commonality of lip rounding across modalities.

More in-depth treatment of the rationale and evidence for the general direct realist approach to speech perception can be found in other reports (e.g., Best, 1984, 1993, in press-a, in press-b; Fowler, 1986, 1989, 1991; Fowler et al., 1990; Verbrugge, Rakerd, Fitch, Tuller, & Fowler, 1984). I am concerned here specifically with how infants' and adults' perception may be affected differently by experience with the native language, particularly by its phonological structure. What we perceive in both native and nonnative speech appears to depend on what we have learned about the native phonology through experience with that language.

A. Language-Specific Phonetic-Gestural Properties and Perceptual Learning

Recall the basic tenets of perceptual learning according to the ecological perspective: (a) that perceptual systems become attuned by experience to particular types of information, (b) that this involves optimization in the pickup of relevant information, (c) that it entails the discovery of critically distinguishing properties of distal structures and events, and (d) that this is accomplished via perceivers' active search for invariants in the flow of stimulation that most economically specify those crucial properties. Educated attention minimizes uncertainty about objects and events in the world by selecting or extracting reduced information specifically for its ability to critically differentiate things of interest or usefulness to the perceiver. Earlier it was argued that the identity of objects and events is specified by structural and transformational invariants available in the flow of stimulation over time and space. Moreover, recognition of similarities and differences among things often depends on abstraction of higher order invariants that depend on prior detection of other, lower order invariants. As Eleanor Gibson remarked, the critical invariants are generally relational in nature rather than isolated, independent attributes.

To consider how higher order relational invariants might be discovered in speech through perceptual learning, I turn briefly to some central concepts developed in work on an ecological approach to the formation of complex coordinated skills and behaviors (e.g., Kugler, Kelso, & Turvey, 1982; Saltzman & Kelso, 1987; Turvey, 1980, 1990), including speech (Saltzman & Munhall, 1989). The goal of coordination is to maximize the adaptability and flexibility of achieving some goal of action by minimizing the number of separate dimensions that must be directly controlled. As Turvey (e.g., 1980, 1990) and others have argued, this is accomplished by forming task-specific synergies among muscle groups or *coordinative structures*. To understand this concept, consider an example commonly cited by ecological researchers—the task of a puppeteer and the way that the construction of his or her marionette simplifies the control of its movements. By linking the puppet's limbs with strings to a controller bar, the puppeteer obviates the need to move each joint of each limb separately, instead producing coordinated movements among multiple limbs by a single movement of the controller. By this means, the many degrees of freedom controlling the joints of the separate limbs have become joined together into a coordinative structure with fewer degrees that must be directly controlled. Research on locomotion indicates that coordinative structures account for the coordination of flexion and extension of each leg joint in proper sequence during the swing of each leg, the alternation between the legs, and the postual adjustments required throughout for maintenance of balance. Coordinative structures show task-specific flexibility in that temporary perturbations result in automatic, immediate compensatory adjustments among the coordinated elements so that the general goal is preserved without requiring numerous command decisions about specific elements.

Saltzman and Munhall (1989) provide logical and empirical evidence that in speech coordinative structures accomplish the gestural goal of forming a constriction of a particular degree at a particular vocal tract location by harnessing together the specific articulators in ways that automatically compensate for perturbations and contextual variations. The language-specific gestural phasing patterns of Browman and Goldstein's gestural constellations are examples of higher order coordinative structures in speech. Coordinative structures in motor control can form and re-form and operate as emergent properties of self-organizing systems (see Madore & Freeman, 1987; Prigogine, 1980; Prigogine & Stengers, 1984; Schöner & Kelso, 1988; Turvey, 1980, 1990). Emergent properties of self-organizing systems, including their sensitivity to initial conditions, have been proposed as the basis for the evolution of maximal dispersion among the elements of language-specific phonological inventories (Lindblom, 1992; Lindblom, Krull, & Stark, 1993; Lindblom, MacNeilage, & Studdert-Kennedy, 1983), as well as for the ontogeny of phonological organization in the child

(Mohanan, 1992; Studdert-Kennedy, 1989). The latter proposals point to the importance of viewing the native phonology as an organized system when considering how language-specific experience may affect perception of phonetic patterns that fall outside the native phonological system.

Insights about coordinative structures and self-organizing processes, and about the importance of minimizing the degrees of freedom that must be separately controlled, serve as useful heuristics for thinking about perceptual learning of phonetic and phonological structure in native speech. Indeed they are crucial to an ecological approach to the issue, given the direct realist assumption that speech perception entails the pickup of information about the distal articulatory events that produced the signal. The ecological approach assumes that perceivers actively explore the rich flow of multimodal information in spoken utterances for invariant patterns that are of interest or utility to them. Educated perception should therefore actively seek and extract critical features of the coordinative structures responsible for the gestural organization of native speech. These coordinative structures should include language-specific articulatory gestures and constellations of phasing among gestures at all levels in the language—from traditional segments to syllables, words, prosodic phrases, and so on. The information detected for the language-specific coordinative structures would be higher order invariants, consistent with the principle that an attuned perceptual system optimizes information pickup by extracting a reduced stimulus, one that minimizes the degrees of freedom that describe the events producing the flow of stimulation. Analogous to the coordinative structures that combine articulators into the coordinative structures to produce gestural events, detection of higher order invariants would automatically account for contextual variations such as speaking rate and style, allophonic variation due to phonetic context, speaker differences, and so on. Such invariants allow the perceiver to "hear through" lower order variations that are irrelevant to phonetic coordinative structures in native speech. To illustrate, take the case of a man saying *Bob* normally versus while clenching a pipe in his teeth. Bilabial closure for /b/ involves simultaneous jaw and lip narrowing movements, whereas the "ah" vowel involves jaw opening along with tongue body movement for pharyngeal narrowing. When the pipe is clenched, however, the jaws are held in a fixed, nearly closed position. As a result, the speaker must accomplish the bilabial closure solely with the lips and the vowel gesture solely with the tongue. The lower order articulatory invariants of specific jaw, lip, and tongue positions at specific times would thus differ between the two utterances, which together permit an attentive listener to hear whether the speaker's teeth are clenched. But the higher order *phonological* invariant in both utterances is that bilabial closure occurs at both ends of the utterance and a pharyngeal narrowing occurs between the two closures. Thus, the word *Bob* is perceived in both cases (i.e., the listener "hears through" the lower order differences to detect the phonological structure). The higher order description provides "reduced" information relative to the lower order one by capturing fewer individual degrees of freedom.

The perception of nonnative speech sounds by the native-language-educated attention of mature listeners would certainly be influenced by the perceiver's seeking of familiar higher order invariants. In other words, the flip side of the efficiency of extracting native higher order invariants may be an increase in difficulty of essentially "going back down a notch" to pick up the lower order, and therefore more numerous, gestural details in unfamiliar nonnative categories and contrasts that are irrelevant to critical distinctions among native gestural constellations (for further discussion of implications for second-language learning, see Best, in press-b; cf. Flege, in press).

Although language-specific higher order invariants are present in native speech, reflecting the coordinative structure among the distal articulatory events that produced it, most or all of these are initially beyond the perceptual reach of infants. They must still discover how the lower order invariants of the simple articulatory components of gestures, which they are able to detect from early on, are harnessed into higher order coordinative structures or gestural constellations by native speakers. Perceptual learning of the critical relational properties of higher order structural and transformational invariants in native speech should thus entail a progressive reduction in the quantity of stimulus detail that must be detected, analogous to the reduction in directly controlled degrees of freedom that results from the formation of coordinative structures in motor skill acquisition (or coordinated control of marionette limbs). This occurs because infants actively explore utterances to discover the optimal sets of gestural invariants that specify the native language structures that are interesting and useful to them. The latter, of course, continue to change as the infant develops, with the discovery of lower order invariants permitting the further discovery of higher order ones.

By this ecological account, then, to learn to perceive the sound pattern of the native language, that is, its phonological structure, is to discover the critical invariants specifying the various nested levels of gestural constellations in native speech. Learning to detect the crucial higher order invariants means, of course, that there will be developmental change in the perception of native speech categories and contrasts. But given the presumed ability to detect lower order articulatory invariants early on, developmental change in the perception of native patterns may be apparent mainly as increased efficiency in extraction of critical invariants. This increased efficiency may foster the infant's emerging ability to recognize words—sound-meaning relations—by the third quarter of the first year. That is, the infant should more easily and rapidly recognize the crucial gestural properties that define a given word irrespective of the irrelevant variation in its specific details when it occurs in different speech contexts, is produced by different speakers, and so on. But perceptual learning of native gestural constellations also carries implications for developmental change in perception of nonnative phonetic patterns during the same period. Developmental changes in perception of nonnative sounds should be, and are, more dramatic because when the infant begins to discover language-specific invariants in native speech, he or she will detect them in native speech but will often be unable to find those familiar invariants in nonnative utterances.

We turn now to the Perceptual Assimilation Model (PAM), which I developed to account for the developmentally changing effect of experience with a particular language on the perception of nonnative phonetic contrasts (Best, 1993, in press-a, in press-b; Best, McRoberts, & Sithole, 1988; Best & Strange, 1992). I began developing this model several years ago in an attempt to provide a coherent theoretical account for a number of observations in the literature on adult cross-language speech perception and on developmental changes in infant speech perception. Specifically, as indicated at the beginning of the chapter, adults often have difficulty discriminating nonnative phonetic contrasts, whereas young infants have no such difficulty. Before the end of the first year, however, infants also begin to display difficulties discriminating nonnative contrasts. However, no existing theoretical treatment offered a single, comprehensive explanation for (a) why, exactly, language-specific effects might occur in either adults or infants, (b) whether and why the effects might differ between adults and older infants, and (c) what the effects might suggest about the influence of phonological knowledge on perception. Certain complexities in reported adult findings would also have to be accounted for: discrimination levels appear to vary among different types of nonnative contrasts; perception of nonnative contrasts can be improved somewhat through perceptual training or through second language learning, but this also depends on the type of contrast involved; and discrimination of nonnative contrasts can be strongly affected by various task manipulations (the findings are reviewed and discussed in greater detail in Best, 1993, in press-a, in press-b).

Based on the considerations laid out in the preceding portion of this chapter, I used the ecological theory of perception as the foundation for developing a coherent theoretical account of the observations on cross-language speech perception in adults and infants. Thus, PAM is based on the ecologically motivated assumption that efficient detection of native gestural patterns in speech may guide and constrain listeners' pickup of information in nonnative phonetic categories and contrasts. This model is unique in several respects. First, it follows an ecological line of reasoning about perceptual learning rather than relying on innate linguistic abilities, information-processing concepts, or cognitive development. Second, it attempts to provide a unified account for both adult cross-language perception findings and developmental changes in infancy. Third, it is the first to provide a detailed, coherent basis for predicting which nonnative contrasts should be difficult to discriminate and which should be easy, and why. To the extent that PAM is compelling and is able to coherently account for the phenomena of cross-language speech perception in adults and infants, it obligates us to give serious consideration to the ecological approach.

We turn next to an overview of how PAM accounts for the perception of nonnative phonetic patterns by adults. For readers who are familiar with PAM, I should point out that there are several new features, by comparison with earlier versions of the model (i.e., Best et al., 1988; Best, in press-a). Specifically, the relation between assimilation of nonnative segments and discrimination of nonnative contrasts has been clarified, additional discrimination types are now recognized and described, and the developmental aspects of the model are more fully delineated.

VII. PERCEPTUAL ASSIMILATION MODEL (PAM)

The basic premise of the Perceptual Assimilation Model (PAM) is that adults actively seek higher order invariants in speech that specify familiar gestural constellations, whether confronted with native or nonnative utterances. Therefore, what they will perceive in nonnative speech, at least initially when they have had little or no linguistic experience with the language involved, are the similarities and dissimilarities between the nonnative gestural patterns and the familiar gestural constellations of their native language's phonological system (for more traditional accounts of the related phenomena of code-switching and loan-word phonology, see Elman, Diehl, & Buchwald, 1977; Silverman, 1992). For nonnative phonetic patterns whose gestural organization is reasonably similar to the gestural invariant for one or more native phonetic categories, the adult listener is likely to detect native gestural invariants, and the nonnative sound will be perceptually assimilated to the most similar native category(s). At the same time, however, listeners should also detect certain discrepancies between nonnative phonetic patterns and native gestural constellations. After all, they are quite sensitive at detecting foreign-accented utterances of their native language (Flege, 1984; Flege & Fletcher, 1992) and nonnative dialect accents.

Note that these predictions are quite open to the possibility of individual differences among listeners regarding which invariants and discrepancies are detected, and how readily.[7] This is because nonnative gestural constellations are not, of course, exactly the same as the native constellations but only resemble them more or less, that is, they display similarity relations rather than identity relations. The resemblances are generally only partial; indeed, a given nonnative gestural pattern may resemble more than one native constellation. Perception of the cross-language similarities would thus ride on selective attention, which is dependent on the listener's history of perceptual learning with the native language—for example, the particular invariants one learns could vary with the style and breadth of native utterances with which one has been engaged—as well as with other languages or other dialects of the native language (e.g., Chambers, 1992).

For consideration of the possible ways in which listeners may perceive nonnative phonetic patterns, it is useful to conceptualize the *native phonetic domain* as the range of vocal tract sounds that are globally speechlike in their gestural properties, vis-à-vis the types of gestures and constellations employed in the native inventory of phonetic categories (for further development of this concept, see Best, in press-b). Outside of this domain, in nonphonetic space, are vocal tract-generated sounds such as coughs, chokes, laughs, whistles, razzes ("raspberries"), tongue clucking, squeals, and so on. The latter three, and other nonspeechlike vocalizations, occur in infant babbling and sound play. However, many infant vocalizations seem at least globally speechlike to adults; some sound quite similar to native categories (as in /baba/ or /didi/), whereas others sound foreign, not falling clearly in any particular native category (e.g., for an English speaker, the latter might include guttural sounds, tongue trills, etc.; Oller, 1980; Oller & Lynch, 1992; Stark, 1980).

Analogously, there are three broad ways in which a nonnative phonetic segment may be perceived with respect to the native phonetic domain (see Table 1). First, the perceiver may detect some resemblance to the gestural invariant of a native category (or perhaps more than one), in which case the nonnative sound is perceptually assimilated to the native category, that is, categorizable. In cases of assimilation to a native category, the nonnative segment may be virtually identical to the native gestural constellation, such that no cross-language discrepancy is perceived. Alternatively, the nonnative segment may be somewhat discrepant but still sufficiently similar to be perceived as a good or acceptable exemplar of the native category. Or it may be even more obviously discrepant and thus be perceived as a poor exemplar of the category. Second, the nonnative segment may be perceived as

[7]For multilingual listeners, there may also be diachronic variations associated with code-switching, i.e., shifting from use of one language to another may effect changes in which gestural invariants are detected in an unfamiliar phonetic pattern (e.g., Elman et al., 1977; Williams, 1977).

TABLE 1
Perceptual Assimilation of Nonnative Phonetic Segments

1. *assimilated to a native phonetic category*
   a. *identical to native gestural invariant:*
      native sound
   b. *reasonably similar to native invariant:*
      acceptable exemplar of native category
   c. *somewhat similar to native invariant, but noticeable discrepancies:*
      deviant exemplar of native category
2. *falls in unfamiliar region of native phonetic domain, outside any native categories:*
   unclassifiable speech sound
3. *falls in nonphonetic space, beyond the boundaries of the native phonetic domain*
   nonspeech sound

globally speechlike, but its gestural organization may not resemble any particular category in the native inventory very clearly. In this case, it will be perceived as speechlike but will not be assimilated to a specific native category. Rather, it will fall in an unfamiliar area of the phonetic domain and be an uncategorizable speech sound, as are the foreign-sounding elements in infant babbling. Third, the nonnative segment may fall entirely outside the gestural range of the native phonetic domain and thus fail to be assimilated as speech, falling instead in nonphonetic space. These segments are nonassimilable as speech and so will be perceived as nonspeech events, for example, as nonspeech mouth sounds, snaps, clicks, and so on.

However, the assimilation of individual nonnative segments with respect to categories in the native inventory only touches the surface of the phonological component of the listener's language-specific grammar. Phonology encompasses the systematic functional relations among phonetic forms within a language, including distinctive segmental contrasts, allophonic alternations, phonotactic constraints, and other phonological processes (e.g., Jakobson & Halle, 1957; Silverman, 1992). From the ecological perspective of perceptual learning, the invariants that determine category membership differ qualitatively from the higher order relational invariants that capture the *critical* differences that define the systematic relationships among categories. Thus, perceiving category membership can be more basic than recognizing critically distinctive relationships between categories. That is, one can recognize a particular instance of /b/ as an exemplar of the /b/ category because it has a complete bilabial closure and concurrent glottal vibration, without necessarily grasping that the critical difference from /d/ is constriction location.

For category membership, the perceiver may begin by extracting a set of lower order properties of category members. But critical comparisons between categories depend on the abstraction of higher order invariants that conjointly acknowledge the similarities that make comparison possible and capture the differences that crucially set the categories apart with respect to some purpose, such as a phonological contrast that serves to differentiate word meanings (J. Gibson, 1979). A critical contrast between events is characterized by *distinctive* features. Distinctive features do not merely list the lower order properties of the individual classes, but rather they capture the relations between classes that remain invariant over contexts and non-identity-changing transformations and thereby define the uniqueness of each class with respect to the other (E. Gibson, 1963). The distinctive higher order invariants that define phonetic contrasts indicate mere "otherness" and cannot be heard independently of a speech segment (E. Gibson & J. Gibson, 1972), for example, location of constriction in the earlier example. Thus, there are more economical rather than category-defining properties, and they optimize information pickup by an experience-attuned perceiver.

For these reasons, the influence of the systematic functional relations within the native phonology should be more readily apparent in perceptual comparisons between contrasting nonnative categories than in a perceptual response to a single nonnative category. As summarized in Table 2, PAM predicts that listeners will easily discriminate between nonnative categories when they can detect in those sounds an invariant that specifies a critical difference or phonological contrast between gestural constellations in the native language (referred to as a *T*wo-*C*ategory assimilation type, or *TC*). They should discriminate moderately well to very well between a nonnative category for which they detect strong similarity to a given native gestural constellation and another nonnative category for which they detect less similarity (or greater discrepancy) to the same native category (Category Goodness difference, or *CG* assimilation type) versus one for which they cannot detect clear similarity to any single native constellation (*U*ncategorized vs. *C*ategorized assimilation type, or *UC*). When both the nonnative categories bear only a global resemblance to the gestural constellations of (native) speech, but do not assimilate clearly into any particular native phonetic category(s), they will be assimilated as uncategorizable speech sounds (both *U*ncategorizable, or *UU*) and will be moderately to fairly difficult to discriminate, depending on whether they bear any remote similarity to any native category(s) and the extent to which any such similarities overlap between the two nonnative sounds. Discrimination should also be very difficult when both members of the nonnative contrast are perceived to fit *within* a gestural constellation for a single native category equally well (*S*ingle *C*ategory assimilation type, or *SC*). The SC case and the CG case actually fall at different points along a single dimension, in that both involve nonnative contrasts whose members are assimilated to a single native category. Thus, to the extent that prototype effects in perception of phonetic categories (i.e., asymmetries in discrimi-

TABLE 2
Assimilation Effects on Discrimination of Nonnative Contrasts

Contrast Assimilation Type	Discrimination Effect
Two-Category (TC)	**excellent discrimination** each nonnative sound is assimilated to a different native category
Category-Goodness Difference (CG)	**moderate to very good discrimination** both nonnative sounds assimilated to the same native category, but they differ in discrepancy from native "ideal" (e.g., one is acceptable and the other is deviant) *can vary in degree of difference as members of native category*
Single-Category (SC)	**poor discrimination** both nonnative sounds assimilated to the same native category, but are equal in fit to the native "ideal" *better discrimination for pairs with poor fit (equally poor) to native category than pairs with good fit (equally good)*
Both Uncategorizable (UU)	**poor to moderate discrimination** both nonnative sounds fall within unfamiliar phonetic space *can vary in their discriminability as uncategorizable speech sounds*
Uncategorized vs. Categorized (UC)	**very good discrimination** one nonnative sound assimilated to a native category, the other falls in unfamiliar phonetic space, outside native categories
Nonassimilable (NA)	**good to very good discrimination** both nonnative categories fall outside of speech domain and are heard as nonspeech sounds *can vary in their discriminability as nonspeech sounds*

nation around good vs. poor exemplars of a category—e.g., Grieser & Kuhl, 1989; see description in next section) are operative in speech perception, they should combine with the SC and CG assimilation patterns to predict better SC discrimination when both nonnative categories are assimilated as poor (nonprototypical) rather than as good exemplars of the native category and to predict CG discrimination asymmetries that reflect greater category generalization (poorer discrimination) around prototypical exemplars than around nonprototypical exemplars of the native category. Discrimination should be moderate to very good, comparable to the CG assimilation type, if both nonnative gestural patterns are perceived to fall outside the native phonetic domain altogether, in nonphonetic space (*N*on-*A*ssimilated type, or *NA*).

The earlier comparisons of gestural scores for English and non-English phonetic categories illustrate some of these cross-language gestural similarities and dissimilarities. In the Hindi [d̪a]-[ɖa] example (Figure 3), the dental versus retroflex constriction locations do not distinguish English stop consonants; in fact, they occur as phonologically equivalent (i.e., nondistinctive allophonic variants of (alveolar) /d/. As for the Zulu [kʰ]-[k'] example (Figure 4), a distinctive property of the voiceless velar stop in English [kʰ] is a glottal opening gesture coordinated with closure (as in Zulu [kʰ]). This critical gesture is lacking from Zulu [k'], which instead has a glottal closure and is therefore notably discrepant from [kʰ]. The Zulu voiced–voiceless lateral fricatives differ by essentially the same glottal voicing distinction (open glottis versus critically closed glottis) found in similar English fricative contrasts (e.g., /s/–/z/, "sh"–"zh"). Lastly, the dual alveolar + velar closures and the suction-release gesture for Zulu alveolar versus lateral clicks are globally unlike anything in English phonology and resemble nonspeech events such as cork popping and finger snapping rather than being even generically speechlike for most English listeners.

PAM thus predicts that adults' attunement for detecting the articulatory gestural invariants that specify familiar phonetic categories of the native language will foster detection of both similarities and dissimilarities between nonnative segments and the native inventory. Even more importantly for questions about perceptual influences of the native phonological system, discrimination of nonnative contrasts is predicted to depend on the listener's abstraction of higher order invariants that specify distinctive oppositions in the native phonology, as well as on their detection of discrepancies between the native contrasts and gestural properties of contrasting nonnative segments. But what of young infants, who are not yet perceptually attuned to native phonetic categories, and especially to the native phonological system? When and how do infants begin to extract the gestural invariants of native categories and the higher order invariants of critical distinctions found in native contrasts? And how does this early perceptual learning of the phonetic categories and relationships of the native language begin to affect perception of nonnative phonetic forms?

To provide a basis for discussing these issues, we begin with a brief review of empirical findings on developmental changes in infants' perception of native and nonnative phonetic contrasts. Following that, we outline a perceptual learning account of development that appears to accommodate those facts. That outline provides the background for studies I have conducted with students and colleagues to test several predictions of PAM for perception of varying nonnative phonetic contrasts by adults and infants.

A. Developmental Changes in Infant Perception of Phonetic Contrasts

Young infants, up to about 4 months of age, have had relatively limited experience hearing the native language. Even the language experience they have had generally focuses attention more on prosodic patterns than on minimal segmental contrasts. The infant-directed speech that is typically addressed to them is characterized by exaggerated pitch contours and durational properties, relative to adult prosody in most cultures (Fernald et al., 1990; Fernald & Mazzie, 1991; Fernald & Simon, 1984; Grieser & Kuhl, 1988; cf. Bernstein Ratner & Pye, 1984). Moreover, infants from birth to at least 4 months of age prefer listening to infant-directed speech more than to adult-directed speech (Cooper & Aslin, 1990; Fernald, 1984, 1985; Fernald & Kuhl, 1987; Werker & McLeod, 1990). In contrast with its prosodic properties, infant-directed speech is not marked by exaggeration or emphasis of segmental distinctions (Bernstein Ratner, 1984, 1986; Bernstein Ratner & Luberoff, 1984; Malsheen, 1980). Even so, many findings indicate that young infants do discriminate a broad range of consonant and vowel contrasts in nonsense syllables, regardless of whether or not the contrasts occur in their language environment (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Jusczyk & Thompson, 1978; Jusczyk, Copan, & Thompson, 1978; for comprehensive reviews, see, e.g., Aslin, 1987; Aslin, Pisoni, & Jusczyk, 1983; Best, 1984; Jusczyk, in press; Kuhl, 1987). Evidence for developmental decline in discrimination of certain nonnative contrasts is discussed in depth in a subsequent section.

A few phonetic differences have been suggested to pose difficulties for young infants, namely, certain native fricative voicing contrasts (e.g., English /s/–/z/; Eilers, 1977; Eilers & Minifie, 1975) and fricative place contrasts (e.g., /f/–"th" [*think*]; Eilers, Wilson, & Moore, 1977). However, more recent work by those researchers, as well as by others, has shown that infants do discriminate those same contrasts (Eilers, Gavin, & Oller, 1982; Holmberg, Morgan, & Kuhl, 1977; Levitt, Jusczyk, Murray, & Carden, 1988). Moreover, infants discriminate other fricative place of articulation contrasts, both native (e.g., /s/–"sh"; Eilers & Minifie, 1975; Eilers et al., 1977; Kuhl, 1980) and nonnative (e.g., Czech retroflex vs. palatal voiced fricatives; Eilers et al., 1982; Trehub, 1976). The balance of that evidence indicates that young infants can discriminate native and nonnative fricative contrasts.

In addition to the basic discrimination findings, infants under 4 months show other revealing perceptual patterns. When familiarized with a set of syllables that share either a common vowel and different consonants, or the converse, 2-month-olds and newborns can detect the addition of new syllables that differ in either consonant or vowel or both (e.g., Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy, & Mehler, 1988; Jusczyk & Derrah, 1987), even though newborns are more affected by attentional manipulations (Jusczyk, Bertoncini, Bijeljac-Babic, Kennedy, & Mehler, 1990). This pattern suggests that young infants perceive the syllables holistically rather than as a combination of discrete segments. Infants between 2–4 months old can also discriminate three to five syllable utterances whose medial syllables differ, apparently only if the contrasted elements are highlighted by the exaggerated prosodic contours of infant-directed speech or differ on more than one articulatory feature (e.g., /r/–/k/) (Fernald & Kuhl, 1982, cited in Karzon, 1985; Goodsitt, Morse, Ver Hoever, & Cowan, 1984; see review by Jusczyk, 1993).

Vowel prototype or "magnet" effects may also be found quite early. The magnet effect refers to a perceptual pattern in which listeners show preferences for and greater generalization (poorer discrimination) around good rather than poor exemplars of a vowel category (as per adult goodness ratings) (Grieser & Kuhl, 1989). These perceptual asymmetries around good versus poor tokens indicate that perception of vowel categories is not absolute, but rather shows systematic within-category differentiation, an effect that occurs only in humans and not in monkeys (Kuhl, 1991). The discrimination asymmetry for good versus poor tokens has been found in human newborns with both native and nonnative vowels (Walton & Socotch, 1993). By 6 months of age, infants still show the effect for a native vowel (Grieser & Kuhl, 1989) but not for a nonnative one (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Polka & Werker, in press) (the latter findings are discussed in more detail later).

In addition, young infants are able to perceive, for at least some consonants and vowels, an underlying phonetic category identity throughout the variations introduced by different pitch contours, different speakers, and different adjacent segments. Detection of such a phonetic equivalence class would appear as perceptual constancy across such variations in a phonetic category, within the familiarization or background stimuli and within the test stimuli. Perceptual constancy was shown in 1–4-month-olds for discrimination of a vowel contrast presented with pitch contour variations (Kuhl, 1979; Kuhl & Miller, 1982). Similar perceptual constancy in discrimination of a consonant contrast across speaker variations has been found in 2-month-olds (Jusczyk, Pisoni, & Mullennix, 1992), but only if there is no delay between the familiarization and testing phases. Similar memorial effects have been found in adults (Martin, Mullennix, Pisoni, & Summers, 1989). Perceptual constancy across varying phonetic contexts (e.g., /p/ across /pi/, /pɑ/, /pu/; nasalization across /nɑ/, /mɑ/, /ŋɑ/) has been found for both vowels and consonants by 4–6 months of age (e.g., Fodor, Garrett, & Brill, 1975; Hillenbrand, 1983, 1984; Kuhl, 1979, 1980, 1983). Thus far, only native phonetic categories have been tested with infants.

The findings summarized thus far have demonstrated little evidence of developmental changes in basic aspects of infant speech perception for native segmental contrasts, except for some signs of increased susceptibility to attentional manipulations or memorial disruptions in the first 2 months

(Bertoncini et al., 1988; Jusczyk et al., 1992). However, in the final quarter year, there are some clearer indications that perception of native segmental patterns begins to be influenced by experience with the language. As discussed earlier, languages differ in both the inventories of consonants and vowels they employ, as well as in their phonotactic rules regarding permissible sequencing of those elements. When 9-month-olds are permitted to choose between listening to two series of unfamiliar words with English versus Dutch segments and phonotactics, infants from each language preferred listening to the list representing their native language. Younger infants showed no preference between these prosodically similar languages. Although English-learning infants did show a native preference when presented with English versus prosodically different Norwegian, that effect was solely attributable to prosody rather than segmental and phonotactic constraints (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993). The experiential effect on 9-month-olds' preference for segmental patterns is strengthened by recent findings that Dutch infants this age prefer phonotactically permissible versus phonotactically impermissible sequences of Dutch segments (Friederici & Wessels, in press), and that American infants prefer frequently occurring versus infrequently occurring English phonotactic patterns (Jusczyk, Charles-Luce, & Luce, submitted; see also Jusczyk, in press).

Infants' discovery of relations between sound patterns and meaning also begins around the last quarter of the first year, with the beginnings of word comprehension. Infants usually begin producing single words a few months later, at around 12–13 months on average, followed by the emergence of syntactic abilities with their first simple word combinations at around 18 months. A phonetic contrast that young infants discriminated in simple discrimination tests, prior to the emergence of word comprehension, may later be missed altogether as a minimal phonological contrast by the 1-year-old whose comprehension vocabulary still lacks minimal word pairs (e.g., the /d/–/b/ contrast when it appears in *dog* vs. *bog*). This follows from the claim of child phonologists that the earliest linguistic units in the single-word period of child speech are more global than the segment (e.g., Ferguson, 1986; Ferguson & Farwell, 1965; Macken, 1992; Macken & Ferguson, 1983; McCune, 1992; McCune & Vihman, 1987; Menn, 1986; Menn & Matthei, 1992; Vihman, 1992), and that segments are gradually differentiated in both production and perception from these early, more global units (e.g., Goodell & Studdert-Kennedy, 1990; Lindblom et al., 1983; Nittrouer, Studdert-Kennedy, & McGowan, 1989; Studdert-Kennedy, 1986, 1991) due to the pressure exerted by vocabulary expansion on the organization of the lexicon (Lindblom, 1992; Studdert-Kennedy, 1987, 1991). Discrimination of minimal contrasts in meaningful word contexts appears to emerge around 18–19 months of age (Werker & Baldwin, 1991;

see Werker & Pegg, 1992). Similar temporary dips in phonetic ability have also been noted in early word productions, in which they are taken as evidence of progress in the development and systematization of phonological knowledge (e.g., Macken, 1992; Menn & Matthei, 1992).

In the next section, I outline the perceptual learning framework for development of speech perception in infancy (and somewhat beyond). The suggested path of learning is informed, in part, by the findings summarized earlier, in addition to the general principles of the ecological approach to perception. It provides the backdrop for considering the research findings on adults' and infants' perception of nonnative phonetic contrasts, particularly a series of studies motivated by PAM, which is described in the subsequent section.

B. Perceptual Learning and Infant Speech Perception

The basic assumption of the ecological account of perceptual learning offered here is that the type of gestural information the child perceives in speech will change developmentally with increasing attunement to the ambient language. The infant will become better able with experience to detect both finer structure and more encompassing structures in native utterances. Following Eleanor Gibson's (1991) arguments about perceptual learning in general, the detection of gestural patterns in speech should become increasingly specific to the phonological categories and contrasts of the native language; there should be an increasing optimization of attention to them, and pickup of gestural information should become increasingly economical, that is, focus should shift away from irrelevant properties and sharpen for critically distinctive ones. The distinguishing features detected for discrimination should shift developmentally, showing progressive improvement in finding the critical features and in abstracting higher order invariants, both of which reduce the number of comparisons required for discrimination (E. Gibson, 1969, 1971). These are exactly the advantages afforded to an experienced listener by the phonology of the native language. Because the language-specific phonological system reduces lower order phonetic detail to just those distinctive features that are crucial for grammatical purposes (e.g., Archangeli, 1988) and organizes that information into superordinate structures, it allows a sensitized perceiver to take in more information within a given time frame and to minimize uncertainty about the important linguistic units. As experience with the native language optimizes and economizes information pickup, therefore, the infant begins to discover the phonological principles of that language.

This learning will, in turn, be reflected in developmental change in the infant's perception of nonnative categories and contrasts. Progress in perceptual learning about the native language should result in, and be illuminated by, developmental changes in perception of nonnative speech. The suggested pattern of perceptual learning about native phonological structure and its expected effects on infant's perception of nonnative categories and contrasts is summarized in Table 3.

During about the first quarter year of life, at best, very young infants should have attained minimal perceptual learning of the higher order invariants for native segmental contrasts. Their experience with the native language is relatively limited, and the speech typically addressed to them generally focuses attention more on prosodic patterns than on minimal segmental contrasts. The view posited here is that infants initially detect simple differences in low-order articulatory invariants, such as the velar versus alveolar closure location for /g/–/d/, the presence versus absence of a glottal opening gesture for /p/–/b/, or the high versus slightly lower tongue position near the front of the vocal tract for "ee"–"ih." This ability should extend to simple gestural differences in both native and nonnative phonetic contrasts.

Given the assumption that infants detect simple differences in low-order articulatory invariants, it should not be surprising that they can pick up simple gestural commonalities within phonetic categories in their first quarter year, even in the face of certain category-irrelevant variations. That is, they show perceptual constancy for simple phonetic equivalence classes across non-identity-changing transformations. Because lower order articulatory invariants of phonetic categories are not greatly affected by speaker (within a single dialect) and intonation variations but may be affected by phonetic context variations due to coarticulation of consonants and vowels, perceptual constancy across speakers and intonation patterns may be evident earlier in development than perceptual constancy across different phonetic contexts. Thus far, the phonetic constancies demonstrated in the first quarter year (Jusczyk et al., 1992; Kuhl, 1979; Kuhl & Miller, 1982) have involved only speaker and intonation variations. Only the studies with infants in their second quarter year (Fodor et al., 1975; Hillenbrand, 1983, 1984; Kuhl, 1980, 1983) have involved phonetic variations. In addition, given the slower, longer lasting, more global tongue gestures associated with vowels as opposed to the more rapid and localized constriction gestures associated with consonants, perceptual constancy may appear earlier for vowels or may simply be more easily obtained and more robust to attentional manipulations than constancy for consonants. Again, studies of very young infants (Kuhl, 1979; Kuhl & Miller, 1982) have tended to test only vowel constancy, whereas studies with infants in their second quarter year (Fodor et al., 1975; Hillenbrand, 1983, 1984; Kuhl, 1980, 1983) have tested for consonant constancy. The possibility of a vowel versus consonant

TABLE 3.
Perception of Native and Nonnative Contrasts in Infancy and Early Childhood

Developmental Phase	Information Detected	Native Phonetic Categories	Nonnative Phonetic Categories
1st quarter year (0–3 months)	simple articulatory gestures *(language universal)*	discriminates any vowel & consonant difference	same as for native speech
	good vs. poor exemplars of simple gestures *(language universal)*	prototype effects for vowels (and consonants?)	same as for native speech
	invariants of simple gestures under speaker & intonation variations *(language universal)*	perceptual constancy for vowels and consonants	same as for native speech
2nd quarter year (3–6 months)	continues as above *(language universal)*	continues as above	same as for native speech
	invariants of simple gestures under phonetic context variations *(language-specific)*	perceptual constancy for native categories	may fail with nonnative categories
3rd quarter year (6–9 months)	simple relational invariants for vowels *(language-specific)*	discriminates native vowel differences	fails to discriminate nonnative vowels that differ from native relational invariants
	good vs. poor vowels re: relational invariants	prototype effects for native vowel categories	lacks prototype effect for non-native relational invariants

4th quarter year (9–12 months)	simple invariants for native bestural constellations *(language-specific)*	discriminates native vowel and consonant categories	discriminates if able to detect different native invariants, or good vs. poor native invariant or if no speechlike gestures at all
		prefers listening to common native syllable patterns more than nonnative or uncommon native patterns	fails if detects a native invariant but not a goodness difference or if detects speechlike gestures but not any native invariants
Extending to 2nd year (9–17 months)	simple invariants for sound-meaning association	learns to recognize simple native words and meanings re: global gestural patterns	may have difficulty learning meaning associated with nonnative global patterns
18 months	higher order relational invariants for minimal contrast word pairs	detects native phonological contrasts	perception of nonnative phonological contrasts depends on similarity to native contrast invariant
2–5 years	higher order relational invariants among some allophones	tendency toward perceptual equivalence among allophones of a category	no difference in response to nonnative allophones vs. nonnative phonological contrasts
	higher order invariants specifying morphological alternations, etc.		

difference also seems compatible with the findings of Bertoncini et al. (1988) and Jusczyk et al. (1990, 1992). However, further investigation is needed to evaluate both possibilities of early developmental changes in perceptual constancy.

Regardless of these possible stimulus parameter effects on perceptual constancy of phonetic equivalence classes, very young infants should show constancy equally for native and nonnative phonetic categories. To the extent that phonetic categories and contextual effects differ among languages, infants should become attuned to native language patterns, and we should expect to see some language-specific effects emerge later, probably around the second half year. Thus far, however, no studies have examined phonetic perceptual constancy about variations of speaker, intonation, or phonetic context in infants of any age.

The assumption that very young infants detect simple gestural properties of phonetic categories also admits the likelihood that they should also show so-called perceptual magnet effects within the first quarter year, at least for vowels. This is based on the reasoning that prototypes and nonprototypes differ in how well they convey the important gestural properties of a vowel category. This, in turn, would affect how easily perceivers could detect the gestural pattern of the category in the differing stimulus tokens. The notion that there is an articulatory basis for good versus poor vowels is consistent with the quantal theory of speech. The quantal theory demonstrates that certain vowel types are very stable, in that small changes in their articulatory constriction location produce minimal changes in the acoustic pattern of the vowel, whereas other constriction locations are unstable acoustically. Languages tend to avoid the latter locations for possible vowels (Stevens, 1972, 1989). Infants in their first quarter year would be expected to show magnet effects for both native and nonnative vowels, a prediction that is consistent with one recent report (Walton & Socotch, 1993).

Young infants in their first quarter year should not yet recognize the more complex coordination or phasing required for specific native gestural constellations; for example, syllable-initial /l/ in English has an uvular narrowing gesture that follows the tongue-tip closure gesture for /l/, rather than being synchronous with it as in word-final English /l/ and in the Russian "hard" /l/ or as being absent in the Russian "soft" /l/. Only as infants become attuned to detecting invariants for familiar gestural constellations in native speech should they begin to show effects of native language experience on their perception of nonnative contrasts. This sort of native attunement would not be expected until at least the second quarter year (perhaps in perceptual constancy across phonetic context variation) or, more likely, the following quarter year.

By the third quarter year (second half year), infants should progress to discovering and attending to more economical higher order relational invariants found in the native phonology, such as the ratio of the two portions of the vocal tract that fall on either side of the tongue constriction location for a given native vowel. These discoveries are assumed to proceed systematically from less to more encompassing and more economical invariants. Thus, the first sorts of native relational invariants infants are likely to discover are relatively simple ones, such as the ratio between the length of the vocal tract that lies before versus behind the high front tongue constriction for the vowel /i/ ("ee"). Once they detect such invariants, they should begin to show language-specific influences on perception of nonnative vowel contrasts and prototypes. These older infants' abilities to discriminate nonnative vowels and to perceive nonnative vowel prototypes will depend on whether they can detect in those stimuli the relational invariants that they can now detect in native vowels, that is, whether they "assimilate" the nonnative vowels to native categories. If so, performance will further depend on whether the infant assimilates the nonnative vowels as good exemplars of native category and whether two contrasting nonnative vowels are assimilated to the same native category or to different categories. However, we should not expect infants' assimilations to match those of adults completely because infants' detection of native vowel invariants is surely not as well tuned as that of adults, and the invariants they detect may be somewhat lower order than those of adults.

With further experience, by the last quarter of the first year, infants should also begin to recognize the higher order invariants that specify native gestural constellations for consonants, as well as the broader phonotactic patterns of native syllables. For example, they should begin to recognize the higher order relational invariants that specify consonant gestural constellations in the native language, such as the precise phasing between the bilabial closure and the glottal opening gestures for English /p/ (as opposed to the different phasing for French /p/). At this point, infants' listening preferences and discrimination abilities will reflect language-specific influences on perception of nonnative consonants and syllable types (such as phonotactic rules regarding how consonants and vowels may be sequenced to form syllables). Older infants' perception of these sorts of nonnative gestural constellations will also depend on whether and how those patterns provide the higher order gestural invariants they have learned to detect in native consonants and syllable types. Again, these older infants' assimilations of nonnative constellations to native categories is still not expected to match adults' assimilation patterns, which derive from a much more sophisticated level of perceptual learning that incorporates minimal phonological contrast and other even more complex relations among segments in the native phonology.

At this point in development, however, infants would not necessarily perceive allophones of a given phoneme as related variants of a single segment, such as the allophonic relationship among stressed syllable-initial voiceless aspirated /p/ versus unreleased final /p/ versus voiceless unaspirated /p/ after /s/. Instead, they may detect differences among allophones simply as gestural characteristics of differing native syllable patterns. This is because they presumably would not yet have discovered the even higher order invariants that relate allophones to common underlying phonological categories. Such abstract commonalities draw on grammatical relations among lexical items (e.g., different morphological forms of a stem word; see further discussion later), which are still beyond young infants' grasp.

Sound-meaning associations, which relate the higher order gestural constellation of the spoken word to the confluence of contextual signs of its meaning, emerge in comprehension during the final quarter year. Some ecological, perceptual learning accounts of this important discovery have been offered in the literature. For example, parents often repeat a key word several times to their infant under diverse spoken transformations, such as variations in prosody and sentence frame, while they concurrently engage the named object (noun) in different event transformations, such as holding it out or wiggling it back and forth, or while they produce variations on the named action (verb) (Dent, 1990; Dent & Rader, 1979; Goldring (Zukow), 1991; Zukow & Schmidt, 1988). The articulatory gestural component infants extract for such sound-meaning complexes is expected to be less differentiated phonetically than other gestural patterns the same infant might detect in the absence of a sound-meaning relation, because the added dimension of semantic or contextual information for words must be reconciled with the limitations of the infant's perceptual span and the need for economization of information pickup. For this reason, children's early words, in both production and perception, should be differentiated by rather holistic gestural properties and not by the finer grain of minimal contrasts (see Best, in press-a). Minimal contrasts that they discriminated prior to the emergence of meaning are likely to be missed now in sound-meaning complexes. Infants at this point have still not discovered minimal phonological opposition. Discovery of phonological oppositions per se requires detection of finer grained distinctions between the gestural constellations of minimally contrastive, meaningful lexical items. The ability to perceive phonological contrasts as such may not be apparent until the upper edge of the infancy period. Recall that minimal contrasts are part of the phonological component of a language-specific grammar. The perception of minimal contrast in the native language—a minimum requirement of a segmental phonology—should be associated with the so-called spurt in children's productive vocabulary (>50 words), which also predicts the emergence of syntax and morphology (e.g., Macken, 1992). At that point, the comprehension vocabulary, if not also the production vocabulary, should be large enough to include minimally contrastive word pairs such as *bed–bad* or *peas–keys.* To perceive a phonological contrast, a relational invariant must be extracted—the critical segmental distinction that marks a difference in meaning between a minimal pair of words. This characterization is consistent with the earlier summarized finding that older infants begin to detect minimal contrasts in meaningful words around 18–19 months of age (Werker & Baldwin, 1991; see Werker & Pegg, 1992).

Discovery of the still higher order invariants corresponding to numerous other aspects of phonological structure await still more experience with the native language, some probably requiring years. For example, perceptual learning of allophonic relations should depend in part on hearing the same work produced by different speakers and with varying speech styles (e.g., causual, formal, and careful speech), as well as on hearing how morphological operations on words affect the phonetic form of the base word. To illustrate, in American English casual speech /t/ and /d/ have a number of context-conditioned allophonic variants: unreleased stops in final position (e.g., *sit, dad, mad*), rapid tongue taps (flaps) as onsets of noninitial unstressed syllables (*sitting, daddy, kitty*), or glottal stops or nasal-released stops preceding unstressed syllabic /n/ (*kitten* versus *hidden,* respectively). Word pairs that young children are likely to hear could provide them with evidence of some of these phonological relations, as in the unreleased /t/ versus flap in *sit–sitting,* the unreleased final /d/ versus medial flap in *dad–daddy,* the flap versus glottal stop in *kitty–kitten,* and the unreleased /d/ versus nasal release in *hid–hidden.* In these cases morphological transformations of meaningful, known words provide a crucial link among the diverse allophones. Adults may also help clarify some allophonic relations if they "correct" their normal conversational speech patterns by repeating words in careful, precise speech to young children. To illustrate, although they pronounce *kitty* conversationally with a medial flap, they may at times pronounce it carefully for the child (as when correcting the child's spelling errors), with the medial /t/ as a voiceless alveolar stop (see Bernstein Ratner, 1993). An underlying gestural commonality among the diverse allophones of medial /t/–/d/ is apparent in children's productions by 20–22 months (Best, in press-a; Best, Goodell, & Wilkenfeld, in preparation). More abstract phonological relations among allophones may also be highlighted later by learning to read and spell, as is the case for the flapped allophones of /t/ and /d/ (see Treiman, Cassar, & Zukowski, submitted).

Similarly, children may learn about even more abstract phonological relations through frequently used morphological operations. For example, the English voiced–voiceless alternations between /s/–/z/ in noun pluralization (e.g., *cats* versus *dogs*) and between /t/–/d/ in the past-tense forms of regular verbs (e.g., *walked* versus *climbed*) covary with the voicing of the preceding segment. Morphological development during the preschool years (Berko, 1958) should aid children's discovery of related phonological alternations (see also Gerken, Landau, & Remez, 1990; Gerken & McIntosh, 1993). Other structural properties of the native phonological system that may take even longer for the child to fully apprehend in speech include some aspects of linguistic stress and intonation, for which perceptual learning may extend to as late as 7–10 years of age (Cruttenden, 1974).

In the next section, I review recent data from my own and others' laboratories that pertain to the preceding account of perceptual learning about the native phonology and its influence on perception of unfamiliar nonnative phonetic contrasts. The findings are discussed within the framework of the Perceptual Assimilation Model (PAM), although it should be noted that the work of other researchers was generally not motivated by PAM. Although much of the research has involved consonant contrasts, some more recent work focuses on vowel contrasts; these areas are described in separate subsections later. Because PAM's assimilation and discrimination predictions were developed to account for *mature* listeners' perceptions of nonnative phonetic contrasts, adult findings are described first within each area.

VIII. EXPERIMENTAL EVIDENCE ON PAM AND DEVELOPMENT OF PERCEPTUAL LEARNING

A. Consonant Contrasts

PAM predicts that adults' ability to discriminate different nonnative contrasts will vary depending on how they assimilate the nonnative phonetic categories vis-à-vis the phonological inventory of their native language.[8] The assimilation predictions presented here and elsewhere (see Best, 1993, in press-a; Best et al., 1988; Best & Strange, 1992) refer specifically to adults' initial perception of unfamiliar contrasts from languages with which they have had little or no linguistic experience. However, the model could

[8]This claim should also apply to the phonological inventories of other languages for fluent multilinguals who learned their languages during childhood. That is, childhood-onset multilinguals may be able to assimilate unfamiliar nonnnative sounds to categories in any of their multiple languages. Indeed, they may have greater overall sensitivity to the phonetic properties of unfamiliar phonological categories to the extent that early learning of more than one language grants increased recognition of the arbitrariness of linguistic categories. Although this sort of metalinguistic advantage has thus far been argued only for semantic and syntactic knowledge, support has been mixed (e.g., Bialystock, 1988; Rosenblum & Pinker, 1983; see also McLaughlin, 1978).

be extended via the principles of perceptual learning outlined here to account for changes in perception that can occur as adults learn a second language (see Best, in press-b; for an alternative view, see Flege, in press). To review PAM predictions briefly (Tables 1 and 2), adults are expected to show excellent discrimination for nonnative contrasts that are assimilated to two different native categories (TC assimilation type). They should show good to very good discrimination for those that are not assimilated into native phonetic space (i.e., are heard as nonspeech: NA type), or for those assimilated with differing degrees of goodness into a single native category (CG type), or for those in which one pair member is assimilated to a native category but the other is uncategorizable (UC type). Moderate to poor discrimination is expected for nonnative contrasts that fall within unfamiliar phonetic space (i.e., are both heard as uncategorizable speech sounds: UU type), and poor discrimination is expected for those assimilated as equally good exemplars of a single native category (SC type).

Earlier reports of poor discrimination of nonnative consonants by adults have tended to use contrasts that were most likely assimilated as SC types or perhaps as UU types. Discrimination levels for such contrasts should indeed have been low according to the Perceptual Assimilation Model. For example, speakers of Japanese and Korean, who are relatively inexperienced with spoken English, have great difficulty discriminating and differentially labeling English /r/–/l/ (e.g., Gillette, 1980; Goto, 1971; Miyawaki et al., 1975; Mochizuki, 1981; Sheldon & Strange, 1982; Yamada & Tohkura, 1991). Their languages do not have an /l/ category, and their /r/ is not a liquid approximate, as in American English, but rather a flap, more like the medial /d/ in *daddy* (Bloch, 1950; Price, 1981; Vance, 1987). Thus, PAM would expect monolingual Japanese to assimilate both English /r/ and /l/, maybe as poor examplars of their flapped /r/, but more likely as poor exemplars of their approximate /w/ or as uncategorizable speech sounds. The sounds should be rather poorly discriminated by Japanese in any of these cases, although perhaps slighly above chance.

In a study conducted before the development of PAM, Kristine MacKain, Winifred Strange, and I compared American and Japanese listeners' labeling and discrimination of /l/–/r/ in a computer-synthesized continuum ranging from English *rock* to *lock* in acoustically equal steps (MacKain et al., 1981). As expected, the American listeners strongly displayed the phenomenon of categorical perception. That is, they labeled the items at one end of the continuum very consistently as /l/ and the items at the other end as /r/, with a steep category boundary. Correspondingly, their discrimination between items that were three steps apart along the continuum was poor for within-category comparisons but very good for between-category comparisons, with a dramatic peak in discrimination performance at the position of the category boundary found in labeling.

Japanese who had had little English conversational experience, on the other hand, showed nearly flat labeling and discrimination functions, with no category boundary effect and poor discrimination overall. Interestingly, however, a subgroup of Japanese subjects who had had some period of intensive conversational training and/or practice in English showed labeling and discrimination functions similar to those of the Americans, although not quite as high. Thus, the results are compatible with PAM and, in addition, suggest that perceptual of nonnative contrasts can be improved by intensive conversational experience with the language involved (see also Flege, 1989, 1991a; other training approaches may also improve discrimination: e.g., Jamieson & Morosan, 1986; Logan, Lively, & Pisoni, 1991; Pisoni, Aslin, Perey, & Hennessey, 1982; Strange & Dittmann, 1984).

Monolingual English-speaking listeners have also, of course, shown poor discrimination for a number of nonnative contrasts, each of which is most likely to show SC assimilation patterns. For example, Thai voiced versus voiceless unaspirated utterance-initial stops are both good exemplars of English voiced stops and are difficult for English listeners to discriminate (Lisker & Abramson, 1970). Hindi voiceless unaspirated dental versus retroflex stops, which are likely to be heard as /d/, are quite difficult for English listeners to discriminate, as are Nthlakampx (Thompson: Interior Salish) velar versus uvular ejective stops k'/–/q'/, which are likely to be heard as "odd" exemplars of English /k/ (or sometimes as other English sounds; Polka, 1991; Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Lalonde, 1988; Werker & Tees, 1984a). Likewise, the Czech retroflex versus palatal voiced fricatives are poorly discriminated by English listeners (Eilers et al., 1982; Trehub, 1976), who are likely to hear them both as "zh."

Also relevant more generally to the perceptual learning approach are several studies showing that reducing the memory demands of the discrimination task or "stripping away" all acoustic details other than the crucial difference between the contrasting nonnative categories results in increased discrimination of SC-type contrasts (e.g., Carney, Widin, & Viemeister, 1977; Miyawaki et al., 1975; Pruitt, Strange, Polka, & Aguilar, 1990; Werker & Logan, 1985; Werker & Tees, 1984a). Both experimental manipulations reduce the array of information within which the listener must detect the critical differences. With the acoustic manipulation in particular, in reducing or eliminating the irrelevant and redundant stimulus properties, the experimenter both picks out the distinctive features for the listener and simultaneously attenuates the speechlike properties of the stimuli, that is, moves them toward NA assimilation types.

A more comprehensive examination of the Perceptual Assimilation Model, however, requires the comparison of discrimination levels across differing nonnative assimilation types and direct assessment of the listeners' assimilations of the nonnative sounds, that is, their native categories. The first study on this point investigated perception of several click consonant contrasts from Zulu—a southern African Bantu language—by American English adults who were completely inexperienced with any click languages (Best et al., 1988). Clicks should not be assimilable as speech sounds within English phonetic space because their manner and place of articulation are different from anything in the English inventory of gestural constellations. That is, the click contrasts should produce an NA assimilation pattern for most English listeners and should be relatively easily discriminated as nonspeech sounds. Subjects were tested with multiple natural tokens on discrimination of all minimal feature pairings from the 3 × 3 matrix of Zulu click-voicing categories (voiceless, short-lag voiceless, voiceless aspirated) and places of articulation (alveolar, lateral, palatal), which yielded 18 minimal contrasts. According to posttest questionnaires, the listeners assimilated all clicks as various nonspeech sounds (e.g., "a cork popping," "tongue clucks," "finger snaps"), except for one subject who heard some clicks as being similar to English /k/. Performance on an AXB discrimination test was quite good, ranging from 80% correct (chance = 50%) for the most difficult contrast—the alveolar versus lateral voiceless unaspirated pair—to 85%–95% correct for the others. Thus, the PAM prediction of good to very good discrimination for nonnative NA contrasts was met, and performance differed substantially from that reported earlier for nonnative SC (or UU) contrasts.

Several other nonnative assimilation types have been compared in adult studies from my own and other laboratories. In a direct comparison of TC, CG, and SC contrasts, I tested English listeners' discrimination with multiple natural utterances of three additional Zulu contrasts: voiced versus voiceless lateral fricatives, voiceless aspirated versus ejective velar stops /k/–/k'/, and plosive versus implosive bilabial stops. A fourth nonnative pair was the Tigrinya (Ethiopian) bilabial versus alveolar ejective contrast /p'/–/t'/ (Best, 1990). The Zulu lateral fricatives were expected to assimilate to English as TC contrasts, that is, as a voiced–voiceless English fricative contrast involving the tongue tip (i.e., /z/–/s/, "zh"–"sh," or "th" in *this* vs. *think*), perhaps in combination with an /l/. The Tingrinya ejectives were likewise expected to be assimilated as a TC contrast, specifically as "odd" English /p/ and /t/. The aspirated versus ejective velar stops were expected to assimilate as a good versus an "odd" /k/, that is, as a CG contrast. And the plosive versus implosive bilabials were expected to assimilate as nearly equal English /b/s. All PAM predictions were strongly supported. Nearly all subjects assimilated the contrasts as expected, according to a posttest questionnaire that asked them to describe or give English labels to recordings of each nonnative category. Moreover, the levels of AXB discrimination performance were strongly associated with their assimilation patterns. That is, the Zulu and Tingrinya TC contrasts yielded excellent, near-ceiling discrimination. The Zulu CG contrast was discriminated very well, but significantly less well than the TC contrasts. The Zulu SC contrast showed the lowest discrimination, much lower than either the TC or the CG contrasts.

Two other aspects of the results from that study were consistent more generally with perceptual learning principles. First, a recency memory effect was found on the AXB discrimination trials *only* for the SC contrast (plosive-implosive bilabials). Discrimination was significantly better when X matched the B category than when it matched the A category. Second, discrimination performance on all three Zulu contrasts was significantly better for matches on the more English-like pair member. Specifically, Zulu /k/ and /b/ were perceived as more like English /k/ and /b/, respectively, than were the contrasting Zulu /k'/ and implosive bilabial, and the voiceless lateral fricative was perceived as containing an English voiceless fricative (/s/ or "sh") more consistently than the voiced cognate was perceived as containing the corresponding voiced fricative (/z/ or "zh"), even though subjects did assimilate the lateral fricatives as a TC contrast. AXB discrimination was significantly higher when the X was the more English-like /b/, /k/, or voiceless lateral fricative than when it was the less English-like implosive bilabial, /k'/, or voiced lateral fricative.

In another study, which extended the findings of MacKain et al. (1981), we tested several PAM hypotheses by comparing categorical perception in American and Japanese listeners for three related English consonant contrasts that bear differing relations to Japanese phonology (Best & Strange, 1992). The stimuli were computer-synthesized continua for the contrasts /r/–/l/, /r/–/w/, and /w/–/y/. All three are place of articulation contrasts between approximant consonants, involving constriction gestures that are neither complete closures as in stop consonants nor critically narrow as in fricatives. The first is not a phonological contrast in Japanese, as described earlier, and was expected to show SC assimilation or UU assimilation. In the second contrast, /r/ is, of course, nonnative for Japanese, whereas /w/ is a native category but is produced with less lip rounding than in English. Japanese listeners should assimilate this contrast as either a CG difference within the Japanese /w/ category or as a UC contrast with /r/ as an uncategorizable speech sound (or, less likely, as a TC contrast with a very poor Japanese /r/). The /w/–/y/ difference is a phonological contrast in Japanese as in English, although again both elements are pronounced somewhat differently in the two languages. It should therefore be assimilated as a TC contrast by Japanese listeners. Although we did not obtain posttest assimilation judgments from the Japanese listeners, the pattern of consistency in their categorization and discrimination of the three continua fits well with PAM predictions. That is, their best performance was on /w/–/y/, where they matched American listeners' performance levels; their lowest performance was on /r/–/l/, where the Americans performed as well as they did on /w/–/y/ and /w/–/r/. Those Japanese who were least experienced with English showed essentially chance performance levels on /r/–/l/ but were substantially better than chance on /w/–/r/ and especially on /w/–/y/.[9] Japanese with intensive English experience performed more similarly to Americans on /r/–/l/, as summarized earlier for MacKain et al. (1981), and also on /w/–/r/; however, there was no effect of English experience on Japanese performance with /w/–/y/.

Several adult studies from other labs are also consistent with PAM predictions, although they were not designed to test PAM. Werker and Tees (1984a) tested English speakers' discrimination of Hindi breathy voiced versus voiceless aspirated dental stops and dental versus retroflex voiceless unaspirated stops, as well as Nthlakampx velar–uvular ejectives /k'/–/q'/. They found listeners better able to discriminate the first contrast than the other two. This finding is consistent with PAM, given that the latter two contrasts are each likely to be assimilated as an SC contrast, specifically as /d/ and /k/, respectively. The former contrast, however, is likely to be assimilated either as /d/–/t/, a TC voicing contrast, or as a CG difference in which the Hindi breathy voiced dental is heard as a deviant English /t/. The authors had undertaken the study to test whether allophonic experience in the native language may account for variations in discriminability of different nonnative contrasts (see also Werker et al., 1981; Werker & Tees, 1984b). As they noted, although the allophonic explanation may be compatible with good discrimination of the Hindi dental voicing contrast (English has dental /t/ allophones) and poor discrimination of Nthlakampx ejectives (English has no ejective allophones), it is inconsistent with the poor discrimination of the Hindi dental-retroflex contrast (English does have dental allophones of /d/). Interestingly, however, a separate study found that listeners who had had experience with Hindi in their first year of life were better able than those without such experience to discriminate the dental-retroflex contrast as adults (Tees & Werker, 1984).

Two other reports have explicitly evaluated PAM hypotheses against several other possible accounts for variation in perception of differing nonnative speech contrasts. One focused in depth on the Hindi dental-retroflex distinction in initial position, investigating English listeners' perception of that place of articulation contrast within each of four different voicing settings: voiced, voiceless aspirated, breathy voiced (i.e.,

---

[9] In addition, we found that both language groups heard a third, intermediate category between *rock* and *wok*. Tests with a second group of American listeners confirmed our suspicion that this category was clearly heard as an /l/, which falls between /w/ and "y" in place of articulation. See Best and Strange (1992) for further discussion.

voiced aspirated), and voiceless unaspirated (Polka, 1991). The former two voicing patterns occur for initial stops in English, whereas the latter two do not. Performance on the four place-of-articulation contrasts was not uniform, but rather was near chance for the former two voicing patterns, better than chance for the breathy voiced one, and better still for the voiceless unaspirated one.[10] This pattern of results led Polka to reject an account based on the lack of phonological status of the dental-retroflex stop contrast in English, as well as an account based on exposure to dental allophones of /t/–/d/ in English. An account in terms of the acoustic salience of the formant transitions in the various contrasts was also inconsistent with the observed performance pattern, given that formant transitions are most salient acoustically in the voiced dental-retroflex contrast, which was the most difficult for English listeners to discriminate. However, an assimilation account seemed to work well, in that most listeners heard both members of the poorly discriminated voiced dental-retroflex contrast as /d/ and both members of the voiceless aspirated dental-retroflex contrast as /t/, that is, as SC contrasts. But they heard the more easily discriminated voiceless unaspirated dental-retroflex contrast as "th" (*this*)–/d/ and the breathy voiced dental-retroflex contrast as /d/–/t/, that is, the latter two contrasts appear to have been heard as TC contrasts.

In a related study, Polka (1992) examined English and Farsi listeners' perception of the velar-uvular stop distinction in two voicing contexts: voiced (native to Farsi only) and ejective (native to neither language). On the voiced velar-uvular contrast, English listeners perceived the uvular category as "bad" exemplars of English /g/ or as no clear English consonant, thus assimilating the contrast as a CG or UC difference, which they discriminated above chance. Most listeners in both groups performed poorly on the nonnative ejective contrast, describing it either in terms corresponding to an SC assimilation pattern or a UU assimilation pattern. The few subjects in both groups who showed good discrimination described the latter sounds in terms corresponding to TC, CG, or UC assimilation. A separate group of English listeners showed comparable, above-chance discrimination levels on the voiced and the ejective contrast, although with a trend toward better discrimination of the Farsi voiced contrast. They described the Farsi voiced contrast in CG or UC assimilation terms and the ejective contrast in SC or UC assimilation terms. Thus, the findings from these two studies are also generally consistent with the predictions of the Perceptual Assimilation Model.

In contrast with the evidence that adults assimilate nonnative contrasts

---

[10] It should be noted that Polka used a more sensitive discrimination task, that is, one with lower memory demands, than had Werker and Tees (1984a), which may well account for the discrepancy between the two studies in listeners' difficulty with this particular contrast.

with respect to native phonological categories, young infants show little or no effect of the ambient language on their perception of nonnative consonants up to about 8 months of age. A number of studies have shown, however, that language-specific influences begin to appear by 8–10 months of age and are well established by 10–12 months. But how closely does the 10–12-month-old's discrimination of various nonnative consonant contrasts mirror the pattern found in adults? In other words, are 1-year-olds likely to have discovered the same higher order invariants in native speech contrasts as adults have? Have they yet discovered even that most basic aspect of the phonological component of the grammar—phonological contrast? According to the perceptual learning account of infant speech perception developed here, the answer to the last two questions should be "no."

As with the literature on adult tests of cross-language speech perception, initial reports of a decline by 10–12 months of age in infants' discrimination of nonnative consonants used contrasts that adults from their language community assimilate as SC types. In a conditioned head-turn procedure (see Eilers et al., 1977), Werker and colleagues found that English-learning 6–8-months-old discriminate the Hindi voiceless unaspirated dental-retroflex stops, the Hindi breathy voiced versus voiceless aspirated dental stops, and the Nthlakampx velar-uvular ejectives /k'/–/q'/. Yet, by 10–12 months of age infants have essentially ceased to discriminate the first and third of these (the latter age was not tested on the second contrast). Hindi-learning and Nthlakampx-learning infants, of course, still discriminate their native contrasts by 10–12 months of age (Werker et al., 1981; Werker & Tees, 1984a). Moreover, when presented with a computer-synthesized continuum ranging from /b/ to dental to retroflex stops, 6–8-month-old English-learning infants, 10–12-month-old Hindi infants, and Hindi adults perceive three separate categories, whereas 10–12-month-old English-learning infants and English-speaking adults hear only two categories corresponding to /b/ and /d/ (Werker & Lalonde, 1988).

A recent study from my lab extended PAM directly to infants' perception of additional types of nonnative assimilation types (Best et al., 1990). In this study, 6–8-month-old and 10–12-month-old American English-learning infants each participated in three discrimination tests with nonnative consonant contrasts from Zulu, the same ones that had been used in the adult study summarized earlier (Best, 1990): plosive versus implosive bilabial stops, voiceless aspirated versus ejective velar stops /k/–/k'/, and voiced versus voiceless lateral fricatives. The infants were tested using a conditioned visual fixation habituation procedure (see Best et al., 1988; Horowitz, 1975; Miller, 1983). As summarized earlier, English-speaking adults assimilated the lateral fricatives as a TC contrast, the velars as a CG contrast, and the bilabials as an SC contrast. Their discrimination levels followed the order TC > CG > > SC. In the infant study, the 6–8-month-olds discriminated all three contrasts. The 10–12-month-olds, however, failed to discriminate all three Zulu contrasts, unlike both the younger infants and the adults. The most difficult contrast for them was the lateral fricative distinction. Rather than showing even a small (nonsignificant) fixation increase from the end of habituation to the beginning of the test phase, as they had shown for the other two contrasts, in the lateral fricative test they simply showed a further decline or continuation of habituation.

It is noteworthy that the TC lateral fricative contrast was especially difficult for the 10–12-month-olds, given that, as a TC contrast, it was the easiest of the Zulu contrasts for adults. The older infants' difficulty might be related to the fact that most adults assimilated the lateral fricatives to various consonant clusters, many of which were not phonotactically permissible in initial position in English, such as "zhl" and "shl." In other words, the adults did not find a simple segmental contrast in English, or even a pair of permissible phonotactic sequences, to which they could assimilate the lateral fricatives. Not surprisingly, then, the older infants may have been unable to consistently detect any familiar native gestural constellations in the lateral fricatives, and may have instead perceived them as a UU assimilation type, for which discrimination is expected to be poor or perhaps as an SC assimilation type re: English (both Zulu fricatives had /l/-like properties according to many adults.[11] The older infants also failed to show significant discrimination of the velar voiceless aspirated versus ejective /k/–/k'/, which was a fairly easy CG contrast for adults. On this contrast, they showed their largest average increase in fixation during the test phase, which was nearly as large as that of the 6–8-month-olds, but they also showed a high degree of variability. This pattern suggests two possibilities that warrant further investigation: (a) the infants may have assimilated /k/–/k'/ as a CG contrast and shown a prototype asymmetry effect (Kuhl et al., 1992; Polka & Werker, in press) in which discrimination depended on whether they habituated to the English-like Zulu /k/ or the nonprototypical /k'/; (b) some of the infants may have assimilated /k/–/k'/ as an SC contrast, failing to hear that the voicing lag in /k/ is aspirated while the lag in /k'/ completely blocks airflow (i.e., is silent), whereas others may have heard the aspiration difference and shown CG assimilation. The first possibility would result in significant test-order effects in discrimination levels, whereas the second would not.

The good discrimination of the lateral fricative and velar voicing contrasts by both 6–8-month-olds and English-speaking adults, yet poor discrimination by 10–12-month-olds, indicates a temporary dip in develop-

[11]This is a new interpretation, which better handles the full array of findings than the preliminary interpretation offered in Best (in press-a).

ment perhaps comparable to those noted earlier in the phonological properties of toddlers' single-word productions and in their perception of minimal contrasts in meaningful words. Thus, it may be evidence of progress in the discovery of higher order phonological category information in speech. To examine the time course of the transitional period for these two contrasts, Glendessa Insabella and I tested English-speaking 4-year-olds, using the same conditioned fixation habituation procedure as we had with the infants (although they had to be instructed that their fixations controlled the audio and that they should tell us afterward whether "the sounds changed" at some point during the test; Insabella & Best, 1990). We had to assure that this procedure was sensitive enough to detect discrimination for a contrast we knew they should be able to hear, so all children had to show fixation recovery on one test with English /b/–/d/. Because these older children would only tolerate two tests in a session, we gave one group the Zulu lateral fricative distinction as their second test; the other group got the Zulu velar-voicing contrast as their second test. The 4-year-olds, unlike the 10–12-month-olds, easily discriminated the /k/–/k'/ contrast. However, they still failed to discriminate the lateral fricative contrast. Thus, they had already come into line with adult performance on the CG contrast, but they still showed depressed performance on the TC contrast that had proven to be easiest of all for the adults. The reversal of the developmental dip for the CG contrast but not for the TC contrast should not be particularly surprising, given the complexity of the adults' assimilation patterns for the latter contrast, as noted earlier. The prolonged difficulty with the lateral fricative contrast is to be expected according to the outline of perceptual learning discussed earlier, in that the most common assimilations for adults involved consonant clusters rather than single segments, and many of the clusters were not even permissible in the initial position in English. However, adults' assimilations for /k/–/k'/ displayed much simpler category goodness differences for a single English segment (/k/).

It is crucial to note, in light of the preceding discussion, that 10–12-month-olds do *not* fail with all nonnative contrasts. In a follow-up study with 6–8- and 10–12-month-olds, using the visual fixation habituation procedure but with a more stringent habituation criterion, infants completed three tests: the Zulu lateral fricatives, the Tigrinya ejective contrast /p'/–/t'/ that adults had assimilated as a TC contrast and discriminated quite well (Best, 1990), and an English fricative voicing contrast (/s/–/z/) (Best, 1991). The younger infants discriminated all three contrasts. This time the older group discriminated an adult TC contrast—the Tigrinya ejectives. Yet they still failed with the TC lateral fricative contrast. This failure could not be attributed to a general difficulty with fricative voicing distinctions, because they were able to discriminate well the native English /s/–/z/ contrast.

Given that they could discriminate the TC ejective /p'/–/t'/ contrast that showed consistent, single-segment-based assimilation by adults, these findings lend strength to the interpretation given earlier for the difficulties 10–12-month-olds and even 4-year-olds have with the lateral fricatives.

Another study showed that older infants also clearly discriminated a nonnative contrast that adults assimilate as an NA distinction, as predicted by PAM in concert with the perceptual learning approach (this was actually the first PAM study in chronological terms). Infants at 6–8, 8–10, 10–12, and also 12–14 months were tested on the Zulu click contrast on which adults had shown their "lowest" discrimination performance—still fairly high at 80% correct—on the lateral versus apical voiceless unaspirated clicks (Best et al., 1988). This study used the same conditioned fixation procedure as Best (1990). All infants also completed a test with English /b/–/d/. All four age groups clearly discriminated the click contrast, even though they could not have had even allophonic experience with such sounds in English utterances. Because we had used a rather different procedure than the head-turn procedure that Werker used in her earlier reports of a decline in 10–12-month-olds' discrimination of several nonnative consonant contrasts (e.g., Werker et al., 1981; Werker & Tees, 1984a), we conducted a follow-up study. Using our fixation procedure, we gave 6–8- and 10–12-month-olds a test on the clicks, one on /b/–/d/ and one on the Nthlakampx velar–uvular ejective contrast /k'/–/q'/ used by Werker (Best & McRoberts, 1989). The procedural difference did not matter—Werker's findings of discrimination at 6–8 months and failure at 10–12 months for the /k'/–/q'/ contrast was replicated, as was our previous finding of continued discrimination for the Zulu clicks at both ages.

All told, then, the infant findings with nonnative consonants suggest increasing sensitivity to native gestural constellations, which negatively influences 10–12-month-olds' perception of many but not all nonnative contrasts. However, the patterning for which nonnative contrasts are discriminated or not by older infants, differs in some telling ways from that of adults in their language community. Although they discriminate two contrasts that adults discriminate fairly easily to very easily—an NA contrast and a TC contrast that adults consistently assimilate to a simple segmental contrast in the native phonology—these older infants fail to discriminate two other contrasts that adults also discriminate quite easily—a CG contrast and another TC contrast that shows a more complex and somewhat idiosyncratic assimilation pattern. These findings are consistent with the possibility that 1-year-olds do not recognize the higher order gestural invariants specifying *phonological relations,* including minimal phonological contrasts. The infant's detection of the somewhat lower order invariants corresponding to native phonetic categories may not mark the emergence of true segmental phonology. Rather, the infant's detection of phonological contrast per se may be crucially linked to a growing awareness of word-meaning associations (see Lloyd, Werker, & Cohen, 1993), which initially reflect gestural organization at the word or phrase level rather than the segmental level (e.g., Studdert-Kennedy, 1989, 1991). As stated earlier, perception of minimal phonological contrasts in meaningful contexts may not appear until around 18–19 months (Werker & Pegg, 1992), generally coincident with the vocabulary spurt (50+ words) and primitive syntactic constructions in productive language development.

B. Vowel Contrasts

Much less research has examined language-specific effects on adults' or infants' discrimination of vowel contrasts. However, the few available nonnative vowel findings on adults are consistent with PAM predictions, excepting that thus far no vowel contrasts have met the definition of nonassimilable type, that is, none are perceived as nonspeech sounds. The possibility of NA vowel contrasts, in fact, seems quite remote given the basic commonality of voicing and manner of gestures involved in vowel production. Vowels are associated with a more open vocal tract than consonants and slower, more global gestures involving primarily the larger extrinsic muscles rather than the small intrinsic muscles of the tongue (with some concomitant jaw and lip movements; e.g., Fowler, 1980). Vowel color is differentiated primarily by the location and height of the tongue at its closest approximation to the upper surface of the vocal tract. Vowel contrasts may also involve length (duration) and voice quality differences (e.g., creaky voice). Other differences in the production and in the phonological functions of vowels versus consonants may ultimately be important for understanding adult cross-language assimilation patterns and early developmental changes in perception of nonnative contrasts (see Best, 1993). For example, vowels usually provide the sonority peaks in syllable nuclei (open airflow through vocal tract); vowels carry the prosodic properties of utterances much more than consonants do; speech errors occur among vowels or among consonants but never cross between the two classes; and articulatory movements affect the two classes in opposite manners under stress and speech rate variations (see Fowler, 1980).

Findings on English vowel perception by native Spanish-speaking adults (Flege, 1991, in press) fit well within the PAM predictions, although the research was not motivated by the model. Spanish contains only five vowels: /i/ as in *si*, /a/ as in *casa* (more fronted than English /ɑ/), /e/ as in *mes* (roughly "ay" but not diphthongized as in English), /o/ as in *yo* (not diphthongized as in English), and /u/ as in *su* (not diphthongized as in English). It does not have "eh," "ih," /æ=/ as in *bat*, "uh," short "oo" as in *book*, "aw," or several other English vowels. Thus, English /ɑ/ should be assimilated by Spanish listeners as a moderately deviant exemplar of Spanish /a/. English "ih," "eh," and /æ=/ should be heard as uncategorizable vowels (with respect to each other), or perhaps as poor category exemplars with respect to Spanish /i/, /e/, and /a/, respectively. That is, English "ih"-"eh" and "eh"-/æ=/ should be assimilated as UU types vis-à-vis Spanish phonology, and thus should show relatively poor discrimination, whereas /ɑ/-/æ=/ may show SC or weak CG assimilation pattern and rather poor discrimination. In contrast, /i/-"eh" should show UC or TC assimilation and near-perfect discrimination, whereas "ih"-/i/ should likewise show UC assimilation or a strong CG difference and very good discrimination. Discrimination levels for these contrasts in a recent study by Flege (in press) are consistent with this assimilation account. All contrasts described except for /i/-"eh" were tested with native Spanish listeners. They showed very good discrimination for "ih"-/i/ and poor discrimination for the other three contrasts. The relation between discrimination performance and actual assimilation patterns cannot be determined, however, because the listeners' assimilations were not assessed. Flege accounts for the findings with his Speech Learning Model, which is concerned with whether nonnative sounds are "identical," "similar," or completely "new" with respect to native phonological categories (for details, see Flege, 1991b).

Also compatible with assumptions about adults' assimilation of nonnative segments to their native phonology, Rochet (in press) found differences in the assimilation of the Canadian-French high front-rounded vowel /y/ by Portuguese and English listeners that corresponded to differences in productions of /i/ and /u/ in those two languages. Specifically, English listeners strongly tended to assimilate French /y/ as a /u/, whereas Portuguese listeners assimilated it as an /i/. Also, Polka (submitted) found that English listeners assimilated German high front lip-rounded /y/ and high back-rounded /u/ as a strong CG difference for English short "oo," and German mid-high front-rounded /Y/ versus mid-high back-rounded /U/ as a weaker CG difference for short "oo." She assessed assimilation patterns directly via a keyword identification task, in which listeners had to choose from a list of words that reflected the inventory of English vowels (e.g., *hid, hoed, heed, heard,* etc.) to characterize the perceived closest match for each nonnative vowel. Discrimination was very good for both German contrasts, but significantly better for /y/-/u/ than for /Y/-/U/, which Polka interpreted to be consistent with PAM's predictions.

Finally, in a recent study completed in my laboratory (Best, Faber, & Levitt, in preparation), English-speaking adults were presented with three French vowel contrasts, two Norwegian contrasts, and a Thai contrast. The nonnative vowel contrasts tested were: French high front-rounded /y/ versus mid front-rounded /œ/ were generally assimilated as the TC contrast long versus short "oo" (*boot-book*), and French /œ/ versus less rounded French schwa /ə/ were generally assimilated as the TC short "oo"-"uh." Both were discriminated very well. Similarly, the Norwegian high front in-rounded /ʉ/ and high front unrounded /i/ were assimilated unanimously as the TC contrast short "oo"-"ee" and were also discriminated perfectly. French /o/-/õ/ (nasalized "o") were assimilated as either a strong CG difference for English "o" or as a TC contrast (e.g., "o"-"aw") and were discriminated very well. Thai high back unrounded /ɯ/ and high mid-back unrounded /u/ were assimilated as either a moderate CG difference for English "uh" or, for some subjects, to the TC contrast short "oo"-"uh" and was discriminated slightly less well than the other TC and CG contrasts. Finally, Norwegian high front out-rounded /y/ (which has less lip-rounding than French /y/; Linker, 1985) and /i/ were assimilated by nearly all subjects as comparably good /i/, that is, as a SC type; discrimination was much poorer for this contrast than for the others. When individual subjects' assimilations were grouped according to TC type versus CG type versus SC type, regardless of the specific nonnative vowels involved, the results clearly upheld PAM predictions: Discrimination was near ceiling for TC assimilations, very good but significantly lower for CG assimilations, and much lower for SC assimilations.

Three very recent findings with infants are relevant to understanding the course of perceptual learning for vowels, although only one explicitly evaluated PAM hypotheses. All three studies point to differences between vowels and consonants in the development of native-language effects on perception. In one study of 6-month-olds, English-learning and Swedish-learning infants showed vowel prototype effects only for a native vowel and not for a nonnative one (Kuhl et al., 1992). Comparison of this result to the vowel prototype effects found for both native and nonnative vowels in English- versus Spanish-learning newborns (Walton & Socotch, 1993) suggests a developmental decline between birth and 6 months in detecting goodness-of-fit differences for unfamiliar vowel categories. this suggests that the invariants detected in native vowels by 6-month-olds versus younger infants are different, a possibility supported by a third recent finding. Both German CG vowel contrasts from the Polka (submitted) adult study described earlier were discriminated by 4 1/2-month-olds, who showed no asymmetry in discrimination between the more English-like and the less English-like vowel in each pair. That is, there was no vowel prototype effect on discrimination. However, by 6 months of age, infants discriminated the German vowels only if the habituation or background stimulus was a nonprototype for English (according to the adult judgments), consistent with greater generalization to the prototype than the nonprototype. By 10–12 months, discrimination of both German contrasts failed regardless of the direction of stimulus change (Polka & Werker, in press). The results provide another example of nonnative contrasts that are discriminated quite well as CG contrasts by adults in the infants' language environment, but that are not discriminated by infants over a certain age, the developmental pattern that was found for discrimination of Zulu /k/–/k'/ (Best et al., 1990). Taken together, the infant vowel perception findings suggest that native language effects appear earlier for perceptual prototype effects for nonnative vowels (around 6 months) than for discrimination of nonnative consonant contrasts (around 10–12 months). The argument offered here is that infants discover relational invariants associated with native vowels earlier than higher order invariants associated with native consonants.

Why do infants show changes in perception of nonnative vowels earlier than consonants? Why does the emergence of native-language effects on vowel perception but not consonant perception precede infants' earliest word-meaning associations? Both observations suggest that the invariants infants first discover in native vowels are simpler and/or easier to detect than those discovered in native consonants. There are a number of possible reasons for this developmental asymmetry. Vowel invariants may be easier to discover because the slower vowel gestures are more stable within the flow of information and are evident over a longer period of time than consonants. Different gestural invariants may be extracted for the two classes because the style and complexity of articulatory movements differ. Vowels also carry the prosody of an utterance. Thus the information for vowel invariants may be salient to the young infant at the broader and more attention-getting prosodic level of sound structure in utterances.

IX. FURTHER WORK ON LANGUAGE-SPECIFIC ATTUNEMENT TO SPEECH

Generally, the findings on adults' and infants' perception of nonnative segmental contrasts fit well with the Perceptual Assimilation Model and the basic principles of an ecological approach to perceptual learning of the information in native speech. However, a number of important questions remain unanswered and must be pursued in future research. For example, we still do not know how or even whether infants actually assimilate nonnative sounds to native phonetic categories. Nor do we know which features or invariants they actually extract from either native or nonnative speech. Generating the methodology for assessing these issues will not be easy. Ultimately, techniques will also be needed to investigate the development of perceptual sensitivity to more abstract phonological properties such as allophonic relations, allomorphy (e.g., the voiceless vs. voiced plural marker in *cats* vs. *dogs*), and grammatical effects on phonetic forms (e.g., unreleased /t/ in *sit* vs. flap in *sitting*).

Indeed, it is still largely unknown exactly what information is captured in the invariants for adult speech perception, especially the higher order invariants, although cross-modal speech perception research indicates that the crucial information is gestural in nature and is not specified in purely auditory terms but rather is amodal (e.g., Fowler & Dekle, 1991; Summerfield, 1978; Walton & Bower, in press). Much more work will be needed on this issue, which should benefit from the ecological approach to speech production and its phonological organization (e.g., Browman & Goldstein, 1989, 1990a, 1992a; Kelso et al., 1986; Saltzman & Munhall, 1989). It seems likely that characterizing the invariants in speech perception will depend on careful mathematical and physical analyses as it has in other domains in which, for example, a single parameter (termed *tau*) has been mathematically determined to be the singular invariant that specifies time to contact for an observer moving toward an object (Lee, 1976; Lee, Young, & Rewt, 1992) or for a trajectile moving toward an observer (Savelsbergh, Whiting, & Bootsma, 1991; see also Michaels & Oudejans, 1992), including audible but unseen objects rolling toward a listener (Shaw, McGowan, & Turvey, 1991).

In searching out the higher order invariants for perception of native and nonnative speech, it will probably be necessary also to view the native phonology as an organized system. That is, ultimately it will be important to conceive of the perceptual effects of phonological differences between languages more comprehensively, as effects of systemic differences, and not simply differences in elements or contrasts that one language has and another lacks. This caveat is motivated by proposals that phonological systems are self-organizing and specifically that this leads to maximal dispersion among the elements of language-specific phonological inventories (Lindblom, 1992; Lindblom et al., 1993; Lindblom et al., 1983). But even that work has not addressed how the "optimization of phonetic space" by a language might be expected to affect a listener's perception of particular nonnative contrasts. However, as Lindblom points out (Lindblom et al., 1993), the principle of maximal dispersion would benefit the learning of the native sound system by drastically reducing the size of the phonetic space that must be explored to discover the sound patterning of the ambient language. The relationships among elements in the system would help to illuminate precisely which differences are critical in the language and thereby reduce the information that must be picked up subsequently by the perceiver. The Perceptual Assimilation Model is quite amenable to the conception of the phonological system as an optimization of phonetic space by a given language, but further effort is obviously needed to work out the implications in detail.

X. CONCLUSION

What is innate about the development of the phonological component of a language's grammar? That is, what is it that provides the constraints on acquisition of possible phonological systems? By the ecological reasoning presented in this chapter, the answer is that what is innate—what provides the constraints on phonologies and their development—is the structure and dynamic possibilities of the human vocal tract. To a first approximation, this claim is in line with the underlying assumptions of Chomsky and Halle themselves, whose universal phonetic features were initially based on articulatory concepts. The point on which I disagree with them is their assumption that the constraints are specified innately in the mind. By the ecological view proposed here, the constraints are, instead, literally in the *physical* head, in the vocal tract itself, and in the lawful physical effects that its configuration and movements have on the temporally varying shape of its acoustic product.

Chomsky and Halle (1968) were correct in suggesting that the listener who knows a language hears the phonetic shapes made familiar by experience with that language. This claim, I have argued, can be extended even to predict that the listener hears echoes of those familiar, native phonetic shapes in the nonnative sounds and contrasts of unfamiliar languages. But I part ways with their reasoning about the causal mechanisms and about the source of listeners' knowledge. Instead, I claim that listeners hear the phonological structure of their native language in nonnative speech because they have learned to detect the gestural invariants that are directly available in the information flow from the language environment. Listeners become attuned to these gestural patterns and pick up the invariants specifying those familiar patterns wherever the stimulation provides criterial evidence for them, even in nonnative sounds. This attunement to native gestural invariants begins in infancy but extends over development and into adulthood, where it should even help to account for perceptual changes during the learning of additional languages.

XI. REFERENCES

Abbs, J., & Gracco, V. (1984). Control of complex motor gestures: Orofacial muscle responses to load perturbations of the lip during speech. *Journal of Neurophysiology, 51,* 705–723.

Anderson, S. R. (1985). *Phonology in the twentieth century: Theories of rule and theories of representations.* Chicago: University of Chicago Press.

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology, 5,* 183–207.

Archangeli, D., & Pulleyblank, D. (in press). *The content and structure of phonological representations.* Cambridge, MA: MIT Press.

Berko, J. (1958). The child's learning of English morphology. *Word, 14,* 150–177.

Bernstein Ratner, N. (1984). Patterns of vowel modification in mother-child speech. *Journal of Child Language, 11*, 557-578.

Bernstein Ratner, N. (1986). Durational cues which mark clause boundaries in mother-child speech. *Journal of Phonetics, 14*, 303-309.

Bernstein Ratner, N. (1993). Interactive influences on phonological behavior: A case study. *Journal of Child Language, 20*, 191-197.

Bernstein Ratner, N., & Luberoff, A. (1984). Cues to post-vocalic voicing in mother-child speech. *Journal of Phonetics, 12*, 285-289.

Bernstein Ratner, N., & Pye, C. (1984). Higher pitch in BT is not universal: Acoustic evidence from Quiche Mayan. *Journal of Child Language, 11*, 515-522.

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General, 117*, 21-33.

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

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

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

Best, C. T. (1993). Emergence of language-specific constraints in perception of nonnative speech: A window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 289-304). Dordrecht, the Netherlands: Kluwer.

Best, C. T. (in press-a). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words.* Cambridge, MA: MIT Press.

Best, C. T. (in press-b). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research.* Timonium, MD: York Press.

Best, C. T., Faber, A., & Levitt, A. (in preparation). Association between adults' perceptual assimilation and discrimination of diverse nonnative vowel contrasts.

Best, C. T., Goodell, E., & Wilkenfeld, D. (in preparation). Phonologically-motivated substitutions in a 20-22 month old's imitations of intervocalic alveolar stops.

Best, C. T., & McRoberts, G. (1989, April). *Phonological influences on infants' perception of two nonnative speech contrasts.* Presented at the Society for Research in Child Development, Kansas City, MO.

Best, C. T., McRoberts, G. W., Goodell, E., Womer, J. S., Insabella, G., Kim, P., Klatt, L., Luke, S., & Silver, J. (1990, April). *Infant and adult perception of nonnative speech contrasts differing in relation to the listener's native phonology.* Paper presented at meeting of the International Conference on Infant Studies, Montreal, Canada.

Best, C. T., McRoberts, G. W., & 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.s T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20*, 305-330.

Bialystock, E. (1988). Levels of bilingualism and levels of linguistic awareness. *Developmental Psychology, 24*, 560-567.

Bloch, B. (1950). Studies in colloquial Japanese IV: Phonemics. *Language, 26,* 86–125.

Bohannon, J. N., MacWhinney, B., & Snow, C. (1990). No negative evidence revisited: Beyond learnability or who has what to prove to whom. *Developmental Psychology, 26,* 221–226.

Bornstein, M. H. (1979). Perceptual development: Stability and change in feature perception. In M. H. Bornstein & W. Kessen (Eds.), *Psychological development from infancy* (pp. 37–81). Hillsdale, NJ: Erlbaum.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3,* 219–252.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 62,* 201–251.

Browman, C. P., & Goldstein, L. (1990a). Representation and reality: Physical systems and phonological structure. *Journal of Phonetics, 18,* 411–424.

Browman, C. P., & Goldstein, L. (1990b). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics, 18,* 299–320.

Browman, C. P., & Goldstein, L. (1990c). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 341–376). Cambridge, UK: Cambridge University Press.

Browman, C. P., & Goldstein, L. (1992a). Articulatory phonology: An overview. *Phonetica, 49,* 155–180.

Browman, C. P., & Goldstein, L. (1992b). Response to commentaries. *Phonetica, 49,* 222–234.

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America, 62,* 961–970.

Catford, J. C. (1988). *A practical introduction to phonetics.* Oxford: Clarendon Press.

Chambers, J. K. (1992). Dialect acquisition. *Language, 68,* 673–705.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Chomsky, N. (1972). *Language and mind.* Cambridge, MA: Harcourt Brace Jovanovich.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York: Harper & Row.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (submitted). Do infants perceive word boundaries. An empirical approach to the bootstrapping problem for lexical acquisition.

Clements, G. N. (1985). The geometry of phonological features. *Phonology Yearbook, 2,* 225/N252.

Clements, G. N. (1992). Phonological primes: Features or gestures? *Phonetica, 49,* 181–193.

Cohn, A. (1990). Phonetic and phonological rules of nasalization. *UCLA Working Papers, 76.*

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61,* 1584–1595.

Costa, P., & Mattingly, I. G. (1981). Production and perception of phonetic contrast during phonetic change [Abstract]. *Journal of the Acoustical Society of America, 69,* S67.

Cruttenden, A. (1974). An experiment involving comprehension of intonation in children from 7 to 10. *Journal of Child Language, 1,* 221–231.

Dent, C. H. (1990). An ecological approach to language development: An alternative functionalism. *Developmental Psychobiology, 23,* 679–703.

Dent, C., & Rader, N. (1979). Perception, meaning, and research in semantic development. In P. French (Ed.), *The development of meaning: Pedolinguistic series* (pp. 178–230). Tokyo, Japan: Bunka Hyoron.

Di Paolo, M. (1992). Hypercorrection in response to apparent merger of /ɑ/ and /ɔ/ in Utah English. *Language & Communication, 12,* 267–292.

Di Paolo, M., & Faber, A. (1991). Phonation differences and the phonetic content of the tense-lax contrast in Utah English. *Language Variation and Change, 2,* 155–204.

Kent, R. D., Carney, P. J., & Severeid, L. R. (1974). Velar movement and timing: Evaluation of a model for binary control. *Journal of Speech and Hearing Research, 17*, 175–177.

Krakow, R., Beddor, P. S., Goldstein, L., & Fowler, C. A. (1988). Coarticulatory influences in the perceived height of nasal vowels. *Journal of the Acoustical Society of America, 83*, 1146–1158.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1982). On the control and coordination of naturally developing systems. In J. A. S. Kelso & J. E. Clark (Eds.), *The development of movement control and coordination* (pp. 5–78). Chichester, UK: John Wiley.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowels. *Journal of the Acoustical Society of America, 66*, 1668–179.

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

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development, 6*, 263–285.

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

Kuhl, P. K., & Miller, J. D. (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. *Perception & Psychophysics, 31*, 279–292.

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

Labov, W. (1974). On the use of the present to explain the past. In *Proceedings of the 11th International Congress of Linguists* (pp. 825–852). Bologna: Società Editrice Il Mulino.

Labov, W., Karen, M., & Miller, C. (1991). Near mergers and the suspension of linguistic contrast. *Language Variation and Change, 3*, 33–74.

Labov, W., Yaeger, M., & Steiner, R. C. (1972). *A quantitative study of sound change in progress* (Rep. on NSF contract GS-33287). Philadelphia: U. S. Regional Survey.

Ladefoged, P. (1982). *A course in phonetics*. New York: Harcourt Brace Jovanovich.

Leben, W. (1978). The representation of tone. In V. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 000–000). New York: Academic Press.

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. *Perception, 5*, 437–459.

Lee, D. N., Young, D. S., & Rewt, D. (1992). How do somersaulters land on their feet? *Journal of Experimental Psychology: Human Perception and Performance, 18*, 1195–1202.

Levitt, A., Jusczyk, P., Murray, J., & 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. (1992). The relation of speech to reading and writing. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 000–000). Amsterdam: Elsevier.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1–36.

Lindblom, B. (1992). Phonological units as adaptive emergents of lexical development. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 131–164). Timonium, MD: York Press.

Lindblom, B., Krull, D., & Stark, J. (1993). Phonetic systems and phonological development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 399–409). Dordrecht, the Netherlands: Kluwer Academic.

Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1983). Self-organizing processes and the explanation of phonological universals. In B. Butterworth, B. Comrie, & O. Dahl (Eds.), *Universals workshop* (pp. 181–203). The Hague: Mouton.

Linker, W. (1985). A cross-linguistic study of lip position in vowels. *UCLA Working Papers in Phonetics, 51,* 1–35.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments on comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences.* Prague: Academia.

Lloyd, V. L., Werker, J. F., & Cohen, L. B. (1993, March). *Age changes in infants' ability to associate words with objects.* Paper presented at meeting of Society for Research in Child Development, New Orleans, LA.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89,* 874–886.

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

Macken, M. A. (1992). Where's phonology? In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 249–269). Timonium, MD: York Press.

Macken, M. A. (1993). Developmental changes in the acquisition of phonology. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 435–449). Dordrecht, the Netherlands: Kluwer.

Macken, M., & Ferguson, C. S. (1983). Cognitive aspects of phonological development: Model, evidence, and issues. In K. E. Nelson (Ed.), *Children's language* (pp. 255–282). Hillsdale, NJ: Erlbaum.

Madore, B. F., & Freeman, W. L. (1987). Self-organizing structures. *American Scientist, 75,* 252–259.

Malsheen, B. J. (1980). Two hypotheses for phonetic clarification in the speech of mothers to children. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology* (Vol. 2, pp. 173–184). New York: Academic Press.

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

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

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., & Fei, X. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development, 57* (4, Serial No. 228).

Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 676–684.

McCarthy, (1986). OCP effects: Gemination and anti-gemination. *Linguistic Inquiry, 17,* 207–263.

McCune, L. (1992). First words: A dynamic systems view. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 313–336). Timonium, MD: York Press.

McCune, L., & Vihman, M. (1987). Vocal motor schemes. *Papers and Reports in Child Language Development, 26,* 72–79.

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

McLaughlin, B. (1978). *Second-language acquisition in childhood.* Hillsdale, NJ: Erlbaum.

Menn, L. (1986). Language acquisition, aphasia and phonotactic universals. In F. R. Eckman,

E. A. Moravcsik, & J. R. Wirth (Eds.), *Markedness* (pp. 241-255). New York: Plenum Press.

Menn, L., & Matthei, E. (1992). The "two-lexicon" account of child phonology: Looking back, looking ahead. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 211-247). Timonium, MD: York Press.

Michaels, C. F., & Oudejans, R. R. D. (1992). The optics and actions of catching fly balls: Zeroing out optical acceleration. *Ecological Psychology, 4*, 199-222.

Miller, C. L. (1983). Developmental changes in male-female voice classification by infants. *Infant Behavior and Development, 6*, 313-330.

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

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

Mohanan, K. P. (1986). *The theory of lexical phonology*. Boston: D. Reidel.

Mohanan, K. P. (1992). Emergence of complexity in phonological development. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 635-662). Timonium, MD: York Press.

Morgan, J. L. (1990). Input, innateness, and induction in language acquisition. *Developmental Psychobiology, 23*, 661-678.

Nittrouer, S., Studdert-Kennedy, M., & 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.

Newmeyer, F. J. (1980). *Linguistic theory in America: The first quarter-century of transformational generative grammar*. New York: Academic Press.

Ohala, J. J. (1990). There is no interface between phonology and phonetics: A personal view. *Journal of Phonetics, 18*, 153-171.

Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology, Vol. 1: Production*. New York: Academic Press.

Oller, D. K., & Lynch, M. P. (1992). Infant vocalization and innovations in infraphonology: Toward a broader theory of development and disorders. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 509-536). Timonium, MD: York Press.

Oshika, B. T., Zue, V. W., Weeks, R. V., Neu, H., & Aurbach, J. (1975). The role of phonological rules in speech understanding research. *IEEE Transaction in Acoustics, Speech and Signal Processing, 23*, 104-112.

Pierrehumbert, J. B. (1990). Phonological and phonetic representation. *Journal of Phonetics, 18*, 375-394.

Pierrehumbert, J. B., & Beckman, M. E. (1988). *Japanese tone structure* (Linguistic Inquiry Monograph Series, 15). Cambridge, MA: MIT Press.

Pierrehumbert, J. B., & Pierrehumbert, R. T. (1990). On attributing grammars to dynamical systems. *Journal of Phonetics, 18*, 465-477.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessey, 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.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America, 89*, 2961-2977.

Polka, L. (1992). Characterizing the influence of native experience on adult speech perception. *Perception & Psychophysics, 52*, 37-52.

Polka, L. (submitted). Linguistic influences in adult perception of non-native vowel contrasts.
Polka, L., & Werker, J. F. (in press). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance.*
Price, P. J. (1981). *A cross-linguistic study of flaps in Japanese and in American English.* Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia, PA.
Prigogine, I. (1980). *From being to becoming: Time and complexity in the physical sciences.* San Francisco: W. H. Freeman.
Prigogine, I., & Stengers, I. (1984). *Order out of chaos: Man's new dialogue with nature.* Toronto: Bantam Books.
Prince, A., & Smolensky, P. (1993). *Optimality theory* (Tech. Rep., TR-2). New Brunswick, NJ: Rutgers University Center for Cognitive Science.
Pruitt, J. S., Strange, W., Polka, L., & Aguilar, M. C. (1990, May). *Effects of category knowledge and syllable truncation during auditory training on Americans' discrimination of Hindi retroflex-dental contrasts.* Presented at meeting of the Acoustical Society of America, State College, PA.
Quine, W. V. (1960). *Word and object.* Cambridge, MA: MIT Press.
van Reenen, P. (1982). *Phonetic feature definitions: Their integration into phonology and their relation to speech, a case study of the feature nasal.* Dordrecht: Foris Publications.
Rochet, B. (in press). Cross-language studies of vowel perception and production. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues.* Timonium, MD: York Press.
Romaine, S. (1988). *Pidgin and creole languages.* London: Longman.
Rosenblum, T., & Pinker, S. A. (1983). Word migc revisted: Monolingual and bilingual children's understanding of the word-object relationship. *Child Development, 54,* 773-780.
Sagey, E. C. (1986). On the ill-formedness of crossing association lines. *Linguistic Inquiry, 19,* 109-118.
Saltzman, E. L., & Kelso, J. A. S. (1987). Skilled actions: A task dynamic approach. *Psychological Review, 94,* 84-106.
Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1,* 333-382.
Sampson, G. (1980). *Schools of linguistics.* Stanford, CA: Stanford University Press.
de Saussure, F. (1959). *Course in general linguistics.* New York: McGraw-Hill (Original work published 1916).
Savelsbergh, G. J. P., Whiting, H. T. A., & Bootsma, R. J. (1991). Grasping tau. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 315-322.
Schöner, G., & Kelso, J. A. S. (1988). Dynamic pattern generation in behavioral and neural systems. *Science, 239,* 1513-1520.
Shaw, B. K., McGowan, R. S., & Turvey, M. T. (1991). An acoustic variable specifying time-to-contact. *Ecological Psychology, 3,* 253-261.
Shaw, R., McIntyre, M., & Mace, W. (1974). The role of symmetry in event perception. In R. B. MacLeod & H. L. Pick, Jr. (Eds.), *Perception: Essays in honor of James J. Gibson* (pp. 276-310). Ithaca, NY: Cornell University Press.
Sheldon, A., & 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.
Silverman, D. (1992). Multiple scansions in loanword phonology: Evidence from Cantonese. *Phonology, 9,* 289-328.
Stark, R. E. (1980). Stages of speech development in the first year of life. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology: Vol. 1: Production* (pp. 73-92). New York: Academic Press.
Steriade, D. (1990). Gestures and autosegments: Comments on Browman and Goldstein's

"Gestures in articulatory phonology." In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*. Cambridge, UK: Cambridge University Press.

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David, Jr. & P. B. Denes (Eds.), *Human communication: A unified view* (pp. 51-66). New York: McGraw-Hill.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics, 17*, 3-45.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training in the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics, 36*, 131-145.

Studdert-Kennedy, M. (1986). Development of the speech perceptuomotor system. In B. Lindblom & R. Zetterstrom (Eds.), *Precursors of early speech* (pp. 205-218). New York: Stockton Press.

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

Studdert-Kennedy, M. (1989). The early development of phonological form. In C. von Euler, H. Forssberg, & 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. Schiefelbusch, & M. Studdert-Kennedy (Eds.), *Language acquisition: Biological and behavioral determinants* (pp. 5-28). Hillsdale, NJ: Erlbaum.

Studdert-Kennedy, M. (1993). Some theoretical implications of cross-modal research in speech perception. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 461-466). Dordrecht, the Netherlands: Kluwer.

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

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

Treiman, R., Cassar, M., & Zukowski, A. (submitted). What types of linguistic information do children use in spelling?: The case of flaps.

Turvey, M. T. (1980). Clues from the organization of motor systems. In U. Bellugi & M. Studdert-Kennedy (Eds.), *Signed and spoken language: Biological constraints on linguistic form* (pp. 41-56). Weinheim, FRG: Verlag Chemie.

Turvey, M. T. (1990). Coordination. *American Psychologist, 45*, 938-953.

Vance, T. (1987). *An introduction to Japanese phonology*. Albany, NY: State University of New York Press.

Vihman, M. (1992). Early syllables and the construction of phonology. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 393-422). Timonium, MD: York Press.

Walton, G., & Bower, T. G. R. (1993). A modal representation of speech in infants. *Infant Behavior and Development, 16*, 233-243.

Walton, G., & Socotch, T. (1993, March). *Human newborns shows a "perceptual magnet effect" for native and nonnative language prototypes*. Paper presented at meeting of the Society for Research in Child Development, New Orleans, LA.

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

Werker, J. F. (1991). The ontogeny of speech perception. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 91-110). Hillsdale, NJ: Erlbaum.

Werker, J. F., & Baldwin, D. A. (1991, April). *Speech perception and lexical acquisition*. Paper presented at meeting of the Society for Research in Child Development, Seattle, WA.

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

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

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

Werker, J. F., & McLeod. (1989). Infant preference for both male and female infant directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology, 43*, 230–246.

Werker, J. F., & Pegg. (1992). In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 131–164). Timonium, MD: York Press.

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

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

Whalen, D. H. (1983). Vowel information in postvocalic fricative noises. *Language & Speech, 26*, 91–100.

Williams, L. (1979). The modification of speech perception and production in second language learning. *Perception & Psychophysics, 26*, 95–104.

Yamada, R. A., & Tohkura, Y. (1991). Age effects on acquisition of nonnative phonemes: Perception of English /r/ and /l/ for native speakers of Japanese. In *Proceedings of the 12th International Congress of Phonetic Sciences* (Vol. 4, pp. 450–453). Aix-en-Provence, France: University of Aix Press.

Zukow, P., & Schmidt, C. (1988, April). *Socializing attention: Perceptual bases for language socialization*. Paper presented at meeting of the International Conference on Infant Studies, Washington, DC.

A MODEL OF PHYSICAL REASONING IN INFANCY*

Renée Baillargeon

UNIVERSITY OF ILLINOIS

---

*This research was supported by grants from the Guggenheim Foundation, the University of Illinois Center for Advanced Study, and the National Institute of Child Health and Human Development (HD-21104 and HD-28686). I would like to thank Jerry DeJong, Judy Deloache, and Steve Reznick for helpful comments and suggestions; Elizabeth Cullum for her help with the data analyses; and Andrea Aguiar, Elizabeth Cullum, Lincoln Craton, Julie DeVos, Marcia Graber, Myra Gillespie, Valerie Kolstad, Laura Kotovsky, Amy Needham, Helen Raschke, and the undergraduate assistants at the Infant Cognition Laboratory at the University of Illinois for their help with the data collection. I would also like to thank the parents who kindly agreed to have their infants participate in the research.

Correspondence concerning this chapter should be sent to: Renée Baillargeon, Department of Psychology, University of Illinois, 603 E. Daniel, Champaign, IL 61820.

CONTRIBUTORS TO THIS VOLUME

Renée Baillargeon
Catherine T. Best
Leslie B. Cohen
Eleanor J. Gibson
Bruce MacFarlane Hood
Lisa M. Oakes
Alan Slater
Linda Patla Spear