A direct realist view of crosslanugage speech perception

Chapter · January 1995

CITATIONS
109

READS
1,806

1 author:

Catherine T Best
Western Sydney University
196 PUBLICATIONS **7,025** CITATIONS

SEE PROFILE

Chapter • 6

#996

A Direct Realist View of Cross-Language Speech Perception

Catherine T. Best

This chapter addresses several related issues that are central to understanding language-specific influences on segmental speech perception, and especially experience-related developmental change, from an ecological theoretical perspective (differing theoretical perspectives are presented in other chapters in this volume: Flege; Kuhl and Iverson; Jusczyk, Hohne, and Mandel; Werker). Specifically, it provides a coherent account of the nature of information perceived in speech, of how that information relates to crucial properties of speech production and reveals the phonetic and phonological organization of the listener's language, of the way these factors influence adults' perception of unfamiliar non-native speech sounds and contrasts, and of the developmental course of perception of native and non-native speech. The discussion here examines the following questions. Are the informational primitives for speech perception acoustic/auditory cues, abstract static phonetic features, or dynamic articulatory-gestural patterns? How do the perceptual primitives participate in the phonological organization of a language? How is it, exactly, that experience with the native language influences adults' perception of non-native speech, particularly the perception of similarities and dissimilarities between native and non-native segments? Finally, what information do infants perceive in native and non-native speech, and how does developmental change reflect learning about the native phonological system?

I approach these issues from the direct realist perspective on perception developed by James and Eleanor Gibson and their pro-

ponents, which is also known as the ecological theory of perception (e.g., E. Gibson 1969, 1991; J. Gibson 1966, 1979; Gibson and Gibson 1955, 1972). Following an overview of the direct realist account of speech perception, discussion focuses in particular on a model I have recently developed to account for cross-language speech perception effects in adults and infants (described in more detail in Best 1993, 1994a 1994b; see also Best, McRoberts, and Sithole 1988, Best and Strange 1992). Studies testing the model's predictions about perceptual assimilation, discrimination, and other aspects of perceiving non-native contrasts have been conducted in collaboration with a number of colleagues and students, whose invaluable contributions I acknowledge at the outset: Alice Faber, Carol Fowler, Glenda Insabella, Andrea Levitt, Gerald McRoberts, Nomethemba Sithole (Shepherd), and Winifred Strange[1].

I turn now to the direct realist account of speech perception to set the stage for the subsequent ecologically motivated examination of language-specific influences on perception of non-native segments and contrasts. This approach has strong implications for understanding both the phonetic details and the phonological organization of a language, via an assumed commonality between the primitives of speech perception and those of speech production. In the next section, I briefly review how the direct realist approach compares with other major theoretical views of speech perception. I also discuss the relevance of the ecological view for gaining insights about early language-specific perceptual learning.

DIRECT REALISM AND SPEECH PERCEPTION

What is meant by a direct realist approach to speech perception? This question can be answered most clearly by first briefly reviewing the aspects of the Gibsons' ecological theory of perception and its philosophical foundations that are most relevant to the present discussion (for a broader coverage of the theory, see E. Gibson 1991; J. Gibson 1966, 1979; Reed and Jones 1982). From that context, the direct realist view of speech perception can be described vis à vis other well-known

[1]I am also very grateful to the small army of student research assistants who have worked in my infant lab during the collection of the data described here: Rosemarie LaFleur, Shama Chaiken, Jane Womer, Stephen Luke, Giovanna Perot, Merri Rosen, Eliza Goodell, Sandra Chiang, Gabby Phillips, Laura Klatt, Laura Hampel, Laura Miller, Geoff Tipper, Leslie Turner, David Fleishman, Amy Wolf, Peter Kim, Cindy Nye, Diane Schrage, Pia Marinangeli, Ashley Prince, Jean Silver, Suzanne Margiano, Pam Spiegel, Barbara Diggs, Iris Gutierrez, Alex Feliz, and RitaElena Mangano. The theoretical model described here has also benefited greatly from numerous discussions with colleagues, especially Cathe Browman and Louis Goldstein, Janet Werker, Linda Polka, and Jim Flege.

theoretical accounts of speech perception. Next, I attempt to clear up some common misunderstandings about the assumptions of the ecological perspective on speech. Finally, I turn to the implications that a direct realist approach has for understanding the relation between language-specific phonetic patterns and phonological systems, for understanding the beginnings of phonological development, and especially for understanding developmental changes in perceiving non-native speech sounds and contrasts.

Ecological Theory of Perception

Direct realism, a philosophical point of view regarding the origins of perceptual knowledge, is one of several divergent traditions in the field of epistemology. Its central premise is that in all cases of perception, the perceiver directly apprehends the perceptual object and *does not* merely apprehend a representative or "deputy" from which the object must be inferred. Thus, the direct realist philosophy is contraposed to the indirect or representational philosophical perspective that underlies mainstream contemporary cognitive science. Direct realism does not require that mental events "stand for" the actual objects of perception or intervene between the perceiver and that which is perceived. It is important to note that the modern philosophy of direct realism allows the possibility that objects of direct perceptual awareness may themselves be nonmaterial (e.g., mathematical relations), or even mental (e.g., Gram 1983). However, it assumes, nonetheless, that such immaterial objects are directly apprehended, rather than being inferred or constructed from indirect representations of themselves. In other words, philosophically speaking, materialism is not equated with direct realism nor is mentalism antithetical to it. Both are orthogonal to the central tenet of direct realism, although this fact has often been misunderstood by representationalists. On the other hand, neither does direct realism imply infallible or obligatory perceptual awareness of all properties of material objects that are in the presence of the observer. Fallibility and, indeed, awareness itself are characteristics not of the objects we perceive, but rather of the acts by which we perceive them. Such acts refer to real actions of the perceiver; for example, exploratory movements of the eyes and hands. They do not refer to inferences and indirect mental processes. Given these characterizations, the mere occurrence of hallucinations, mirages, illusions, misperceptions, and individual variations in perception cannot refute the current conceptualization of direct realism.

The ecological theory of perception developed by James Gibson (1966, 1979) and Eleanor Gibson (1969, 1991) has applied the direct realist philosophy to empirical research on perception. Its basic premises are that perceivers gain direct information from the world

about its contents (objects, surfaces, people, and dynamic interactions and relationships among them); that they pick up this information, via integrated perceptual systems, from the temporally and spatially extended flow of stimulation emanating from the world; and that they, therefore, perceive objects, surfaces, and events directly without mediation by inborn knowledge or acquired mental associations. Thus, the ecological view contrasts with both the rationalist and the empiricist traditions in psychological research. Those traditions share the assumptions that perceivers obtain only impoverished and inadequate stimulus data from the world via their sense organs, that these sense data are, therefore, meaningless in themselves, and that, as a result, those data must be cognitively interpreted. The two classic traditions differ from each other primarily on whether they posit that the cognitive injection to the impoverished proximal sensory data comes from innate knowledge (e.g., cognitive processes and modules) or from acquired associations (e.g., memory traces). Current cognitive models typically draw from both mental sources for their indirect representational accounts of perceptual and memorial phenomena. Specific cognitive models differ among themselves primarily in the balance they assume between innate and experiential contributions to mental operations. However, the ecological, direct realist approach rejects all these concerns as derivatives of an untenable premise: that the information perceivers receive from the world is impoverished, ambiguous, proximal sensory data.

The central philosophical problem for both the nativist and empiricist perspectives, according to the ecological perspective, is the paradox at the core of all indirect theories of knowledge. Specifically, if sensory inputs convey inadequate information, then we cannot directly know the world. The notion that we know the world only through inference and interpretation translates to saying that we perceive only what we already know is there to be perceived, and that is circular reasoning (J. Gibson 1979). It is also irreconcilable with evolution through natural selection, which depends on a species' fit to those properties of its niche in the world that are relevant to satisfying its procreative and nutritional needs. If stimulus information were impoverished with respect to real-world events and objects, then the cognitive mechanisms that have been posited for constructing indirect representations of the world could never evolve naturally because the fit of their outputs to events and objects in the world could never be verified.

Because of these and other concerns about indirect theories, the ecological approach to perception argues instead that all animals, for the sake of their survival, *must* know the world directly from information available in stimulation. Stimulation must have extent over

both time and space without having to be mentally "glued together," and it must be picked up actively by perceptual systems rather than simply impinging on isolated sensory organs as instantaneous snapshots or "sound bites." The sensory organs are but one inseparable part of the integrated perceptual systems that have evolved in animals to extract veridical information about distal objects and events from the multimodal flow of stimulation. These perceptual systems are really coordinated perceiving–acting systems for exploration of the world, involving active movements of both the information-detecting organs (eyes, hands, etc.) and of the perceiver as a whole (e.g., locomotion). For example, the eyes do not work in isolation, but are integral elements in a head that is part of a body with hands and feet, all of which can traverse time and space for active exploration of the world. Perceivers actively explore the events and objects that shape the information flux, thereby systematically influencing that flux in ways that provide rich information about the distal sources of stimulation.

Special cognitive mechanisms for handling mental representations and indirect inferencing are not needed, because the flow of stimulation provides a rich and reliable source of direct information about the world, and because perceivers engage their integrated perceptual systems in active exploration of the world. In this view, perceptual learning involves increased attunement for detecting higher-order invariants available in the flow of stimulus information, rather than changes in mental representations and inferential processes.

We next consider the perception of speech, in particular. The discussion is framed around similarities and differences between the direct realist approach and the two other widely known accounts of speech perception, the motor theory of speech perception and the psychoacoustic approach. The comparisons described below are summarized in Table I. (For more in-depth treatments of these models and the debates about them, see Best 1984, 1994b; Diehl and Kluender 1989; Fowler 1986, 1989; and Liberman and Mattingly 1985, 1989.) We begin here by examining the nature of stimulus information that each view assumes to be the foundation, or "input," for speech perception.

Comparative Description of Theoretical Accounts of Speech Perception

The psychoacoustic approach assumes that the basic source of information on which speech perception is based is the proximal stimulus at the auditory periphery—the decomposition of the speech signal into the spectral and temporal characteristics of brief moments in the pressure waveform. That is, it assumes the perceptual primitives in speech to be intrinsically meaningless acoustic cues, such as spectrally limited energy distributions, bursts of noise, silent gaps, and so forth

Table 1. Comparison of psychoacoustic, direct realist, and motor theory, accounts of speech perception and its early development.

	Models		
	Psychoacoustic	Direct realist	Motor theory
Assumptions			
Perceptual primitives	Proximal acoustic cues	Distal articulatory gestures	Speaker's *intended* gestures (neuromotor commands)
Perceptual philosophy	Indirect: information processing or mental representations	Direct pick-up of distal information	Indirect, via motor representations
Perceptual mechanisms	Basic auditory system, aided by cognitive processes	Integrated perceptual systems and their exploratory activities	Specialized phonetic processes of the language module
Specificity re: human speech	General across nonspeech and across other species	General across nonspeech and across other species	Specific to speech and to humans
Relation between perception and production	Not addressed (presumably mediated by cognitive processes)	All perceptual systems integrate perceiving and acting: affordances	Single, specialized module is the source of parity between perception and production
Information infants initially perceive	Nonlinguistic (auditory)	Nonlinguistic (gestural)	Linguistic
Effect of language experience	Formation of traces, templates, prototypes	Perceptual attunement economizes pick-up of native gestural invariants	Native phonetic input tunes the speech module

(e.g., Aslin, Pisoni, and Jusczyk 1983; Diehl and Kluender 1989; Jusczyk 1993). Thus, the psychoacoustic primitives are analogous to the simple, two-dimensional edges, lines, angles, and spatial frequency components that result from the decomposition of proximal retinal images and that typically are offered as the primitives of visual perception. The psychoacoustic stimulus patterns could be characterized as "sound bites," akin to classic "snapshot" descriptions of retinal images.

In contrast, both the current motor theory of speech perception (Liberman and Mattingly 1985, 1989) and the direct realist view (e.g., Best 1984, 1993, 1994a, 1994b; Fowler 1986, 1989, 1991; Studdert-Kennedy 1985, 1986, 1989, 1991) assume that the perceptual primitives are articulatory gestures. The direct realist assumption is that the primitives are the actual gestures produced by the speaker's vocal tract, whereas the motor theory assumes that the perceptual primitives are the *intended* gestures represented in the mind/brain (i.e., the neuromotor command structures that control articulator movements). According to the ecological approach taken by proponents of gestural phonology, gestures refer to the formation of constrictions by diverse

articulators at various positions along the vocal tract (Browman and Goldstein 1986, 1989, 1990a, 1992a). According to the direct realist view, this gestural information is available in speech because its acoustic structure is lawfully shaped, according to the principles of acoustic physics (Fant 1960), by the structure of the vocal tract and its dynamic articulatory transformations (e.g., alveolar closure, velum lowering, glottal opening). Gestural information is present in speech to no less an extent than are pure acoustic features. The direct realist view posits that gestural information is *directly* detected in speech. Gestural information is *not* built up from an analysis of simple acoustic features. The acoustic waveform is regarded simply as an energy medium shaped by, and therefore carrying information about, distal vocal tract gestures. The motor theory similarly rejects the information-processing assumptions of the psychoacoustic approach. However, unlike the direct realist approach, the motor theory assumes that speech perception must be mediated by an innate neural module that synthesizes articulatory gestures.

The three views differ also on how the perceptual primitives become related developmentally to such linguistic units as phonogical elements and morphemes. The psychoacoustic argument is that infants learn to associate combinations of intrinsically meaningless and nonlinguistic acoustic features with linguistic units of both meaning (e.g., morphemes, phrases) and phonological structure (e.g., segments, syllables) (e.g., Jusczyk 1981, 1986, 1993, in press, also Jusczyk et al., this volume). That is to say, through experience, the infant forms auditory memory traces (perhaps stored templates or prototypes) that are paired-associates of abstract linguistic entities.

In contrast, both the direct realist perspective and the motor theory assume that infants must discover which constellations of articulatory gestures are used in their native languages; for example, the temporal phasing between alveolar closure, velar narrowing, and opening along the sides of the tongue for the /l/ at the beginning or end of such American English words as <leg> and <pal>. However, these two views differ in their assumptions about whether the level of information that young infants initially detect in speech is linguistic or nonlinguistic.

In superficial similarity to the psychoacoustic approach, the direct realist approach presumes that infants initially perceive only nonlinguistic information in speech. That is, the distal gestures detected by infants are initially devoid of linguistic relevance for them, analogous to the nonlinguistic nature of distal event information they detect in other environmental sounds. In the case of speech, the infant detects vocal events produced by other people. The child, however, soon begins to discover correspondences between higher-order invariants of relations among gestures, or gestural constellations, and linguistically func-

tioning elements such as morphemes, and so forth, which are specific to the language environment. That is, according to the direct realist approach, speech perception shifts developmentally to a linguistic focus. The direct realist and psychoacoustic views differ, of course, in their assumptions about whether the nonlinguistic information is acoustic or gestural and about whether the developmental shift to a linguistic focus results from the formation of auditory memory traces and cognitive processes, or from perceptual attunement to detect the language-specific, higher-order gestural invariants that are available in the flow of multimodal speech information.

The motor theory differs from both the psychoacoustic and the direct realist perspectives in assuming that the information even young infants perceive in speech is linguistic in nature, and that this is accomplished by mechanisms independent of those used for nonlinguistic auditory perception. The motor theory, like the psychoacoustic view, assumes that nonspeech auditory perception begins with detection of the proximal acoustic cues and does not involve direct detection of distal articulatory events. However, unlike both the psychoacoustic and direct realist views, the motor theory argues that a biologically specialized phonetic module relates the incoming speech signal to abstract phonological units via the neuromotor representations of intended phonetic gestures. On the output side, the phonetic module translates such abstract gestural structures as words into neuromotor commands for producing specific utterances. Thus, the perception and production of speech are directly linked. What happens as a result of developmental experience with a particular language, according to the motor theory, is that the phonetic module becomes tuned to the phonetic inventory of that language.

The direct realist approach assumes instead that infants perceive gestures in speech via an integrated general perceptual system that detects information about distal articulatory events, just as perceptual systems detect distal source information about other sound-producing events. The direct realist view posits that the information detected in either case specifies the distal event that produced the sound, including the structure of the objects and surfaces involved, rather than the proximal pattern of acoustic stimulation at the auditory periphery. What the infant learns about speech as a tool for communicative purposes within a particular language is central to understanding speech perception and its early development, according to the direct realist approach. Just as we tend to perceive physical tools (e.g., a hammer, a pair of pliers) in terms of the potential goal-related actions we may achieve with them (i.e., their affordances: Gibson 1966, 1979), we perceive native speech in terms of its linguistic affordances—communicative goals that can be accomplished with it. In other words, per-

ception and production of speech are inextricably linked. Infants detect information in ambient speech about the articulatory gestures that shaped it, as an integral part of learning to use the vocal tract as a tool for achieving language-specific communicative goals. The higher levels of linguistic structure in speech can only be detected by perceivers who have become attuned to language-specific coordinations of higher-order gestural constellations and referential meanings. This begins as the infant discovers, during the last quarter of the first year of life, that specific words and phrases are repeatedly presented in the context of specific objects and events.

Misunderstandings about the Direct Realist View of Speech Perception

One recurring misconception about the direct realist perspective is that it must imply innate knowledge of the human vocal tract in the perceiver's mind. Underlyingly, this misunderstanding indicates some confusion between the motor theory of speech perception and the direct realist perspective. Although these two theoretical approaches share some assumptions, they are opposed on this particular premise, as well as on several others (see Table I). The motor theory does, indeed, assume a form of innate knowledge about the vocal tract, residing within the specialized language module that it posits. But the direct realist perspective eschews this sort of innate knowledge and rejects the notion that neural and/or cognitive mechanisms are needed to interpret/process/decode/draw inferences from the speech signal; that is, to go beyond the information given directly. The direct realist view, instead, states that the integrated perceptual systems of humans and other animals directly detect adequate, veridical information about distal objects and events from the flow of stimulation that emanates from the world and from the perceiver's active exploration of it. For this reason, proponents of the direct realist view have argued that animals may detect low-level gestural information in speech (see discussions in Best 1994a; Fowler, Best, and McRoberts 1990; cf. Dooling, Best, and Brown 1995), although they apparently do not become attuned to certain higher-order invariants for a specific language's vowels (Kuhl 1991). To the extent that higher-order invariants reflect the human communicative affordances of a native language, which are irrelevant for other species, we would not expect other animals to become attuned to such invariants. Nonhuman animals surely do not have innate knowledge of the human vocal tract. Thus, findings of categorical perception and of perceptual constancy across within-category variations in animals' perception of speech segments have been seen as problems for the motor theory. However, the animal findings are not problematic for the direct realist view, because its assumptions about direct perception of gestural information in speech do not

rely on innate knowledge of the vocal tract, but are compatible, on the other hand, with animals' perception of at least low-level distal gestural information. Note also that the direct realist view rejects the motor theory claim (e.g., Liberman and Mattingly 1985) that listeners perceive speech by reference to their own speech motor control or acoustic output. Instead, it claims only that listeners perceive information about the gestures produced by the speaker, irrespective of whether they themselves could produce such signals (see also Fowler et al. 1990).

Another misinterpretation of the direct realist view of speech perception, particularly regarding cross-language perceptual effects, is that it leaves no room for learning. It is not entirely clear what misconceptions of the theory underlie this peculiar idea, but it is quite at odds with the ecological perspective. Perceptual learning plays a central role in the ecological theory of perception, reflecting the perceiver's increasing, experience-based attunement to detecting higher-order invariants of objects, surfaces, and events. That attunement increases economy in information pick-up, and increases specificity and differentiation of the critically distinctive information characterizing one object or event as different from another (for more extensive discussion, see Best 1994b; E. Gibson 1991). Perhaps the misconception that the direct realist view of speech perception is incompatible with perceptual learning is just another derivative of conflating direct realism with the motor theory postulate of innate vocal tract knowledge. Or perhaps it belies a misconception that direct realism assumes the direct pick-up of real world information to be both obligatory and infallible, that is, that perceivers *must* detect everything that is available in the flow of information and that they cannot be mistaken in anything they perceive. If this were a correct characterization of the theory, then perceptual learning could be difficult to reconcile, and even more certainly the existence of hallucinations and mirages and illusions would be damning counterevidence. However, as summarized earlier, the contemporary philosophical conceptualization of direct realism argues that fallibility, obligation, and comprehensiveness in perception are characteristics of the acts by which we perceive; for example, active exploration of objects and surfaces. They are not characteristics of the things perceived, nor of the directness or indirectness of perception (i.e., of the mind's internal machinations). Hence, faulty or incomplete perception cannot be taken as refutations of the premises of direct realism. It follows that attention and learning are both integral to the theory of direct perception.

To recapitulate the basic assumptions of the direct realist perspective, it adheres to the principle that perceivers directly detect information in the multimodal flow of stimulation that specifies the distal sources of that flow in the world (i.e., objects, surfaces, events).

This is accomplished through active exploration of the distal sources of stimulation, supported by integrated perceptual systems that evolved for detection of veridical information. Perceptual learning involves increasing attunement for detecting higher-order invariants in the stimulus flow, through experience with specific objects and events. The direct realist perspective rejects the basic assumptions of representational approaches to perception and knowledge, including the notions that we receive impoverished inputs from the world through our sensory organs, that abstract cognitive mediation and/or innate knowledge is, therefore, necessary, and that evolution selected for cognitive mechanisms or innate categories of knowledge.

An Ecological View of Language-Specific Phonetics and Phonological Organization

The likelihood that adults and infants perceive gestural information in speech is probably best illustrated by findings on cross-modal speech perception. When conflicting consonants are presented in synchronous audio+visual syllables, listeners perceive a unified phonetic pattern that includes information from both modalities (McGurk and MacDonald 1976). This indicates that the two modalities convey information to the perceiver about a singular event or perceptual object. The common source appears most likely to be articulatory gestural information carried by the two modalities. Attempts to explain the crossmodal effect purely in terms of discrete, binary phonetic features (i.e., without recourse to gestural properties) have failed to account straightforwardly for asymmetries in the crossmodal effect. Specifically, simple phonetic feature representations are identical for audio /aba/+video /aga/ and for audio /aga/+video /aba/; however, in the former case, the unified percept is usually /ada/; whereas, in the latter, it is /abga/ or /agba/. Nor do learned arbitrary associations apparently provide the linkage between modalities. Crossmodal unity also occurs with discrepant audio–tactile presentations of speech (i.e., hand-felt lip movements synchronized with audio), although prior experience with such audio–tactile associations for speech are exceptionally unlikely. On the other hand, there is no crossmodal effect for synchronous audio+written syllables, although literate adults have had extensive experience (indeed, overt training) with the arbitrary associations between speech sounds and their graphemic representations (Fowler and Dekle 1991). Unlike hand-felt lip movements that result directly from speech gestures, printed text is not a physically lawful result of speech gestures but is instead a conventionalized system of arbitrary ciphers.

In further support of the gestural account, young English-learning infants look preferentially at synchronized video repetitions of back-

rounded /u/ rather than at those of unrounded /i/ when listening to audio repetitions of the unfamiliar front-rounded French /y/ (Walton and Bower 1993). Again, associative experience cannot account for this matching preference, which is instead consistent with the notion that the infants detect gestural information in both heard and seen speech.

According to the direct realist view of speech, it is this sort of amodal gestural information that serves as the foundation for the functional linguistic elements of the phonological systems of individual languages. That is, articulatory gestures are the primitives of which phonological elements are comprised; higher-order constellations among gestures serve as the units of linguistic contrast in phonological systems (e.g., the gestural score of a word—the cast of gestures involved, their relative magnitudes, their temporal phasings) (Browman and Goldstein 1986, 1989, 1990a, b, c, 1992a, b). Thus, the direct realist view is that phonetics and phonology are both grounded in the domain of articulatory gestures, but they tap different levels of invariant structure in that domain, perhaps analogous to the microscopic versus macroscopic views of some physical object (Browman and Goldstein 1990a). This is opposed to the representational view of mainstream linguistics, as well as of the psychoacoustic approach to speech perception, which assume two separate informational domains for phonetics and phonology. That is, by the representational account, phonetic details reside in the physical realm (either articulation or acoustics), but phonological structure resides in the cognitive/representational realm (e.g., Pierrehumbert 1990). The problem of interfacing these two orthogonal domains has been the source of much debate and theoretical speculation, running the gamut from classic generative linear phonology (Chomsky and Halle 1968) to modern nonlinear approaches (e.g., Archangeli and Pulleyblank 1994; Clements 1992; Cohn 1990, 1993; Keating 1988, 1990; Pierrehumbert and Pierrehumbert 1990; Zsiga 1993).The representational approach requires that the child mentally reconstruct the phonology of his or her language because the incoming phonetic details are both impoverished and different in kind from phonological structures. However, in direct realist or gestural phonology neither a complicated interface nor mental reconstruction is needed because phonetic details are informationally continuous with phonological structure, and speech provides a rich flow of direct information rather than an impoverished collection of static cues or features (Best 1994b).

Language-specific differences (and even dialect-specific differences) in the phonetic details of a given underlying phonological segment offer some particularly interesting and useful insights into these two divergent views of the phonetics–phonology relation. A core

premise of traditional linear phonology is that the phonetic implementation of phonological representations is universal and nonlinguistic. Coarticulation and other aspects of the phonetic realization of phonological representations are presumed to follow from general physiological and/or mechanical principles. Thus, they should be cross-linguistically universal. That is, phonology is part of the language-specific grammar, whereas phonetic implementation is assumed to be universal and nonlinguistic.

Recent findings indicate, however, that phonetic implementations are often language-specific, not universal (see, e.g., Fourakis and Port 1986, Mohanan 1986). For example, in English stress-initial /p/ the glottal opening gesture for voicelessness reaches its peak simultaneous with the bilabial release (e.g., Löfqvist 1980, Löfqvist and Yoshioka 1984), causing a long aspirated voice-onset time (VOT) interval between release and the onset of vocalization for /a/. However, in French and Spanish (as well as in English unstressed syllables), the glottal opening for /p/ is smaller and occurs before closure release, resulting in a shorter and unaspirated VOT interval (see Goldstein and Browman 1986). The glottal release in Navajo ejective /p'/ is delayed and more forceful compared to Quechua /p'/ (Lindau 1982). Languages differ widely in vowel-lengthening preceding voiced as opposed to voiceless stops (Keating 1984). Even related dialects show such phonetic specificity, as exemplified in the delayed velum-lowering for nasalized vowels in Canadian French as compared to continental French (van Reenen 1982). These sorts of language-specific differences in phonetic implementation cause grave difficulties for the claims of traditional linear phonology. In fact, they complicate the task of mapping from phonological representations to phonetic output even for more modern nonlinear phonological approaches. But the direct realist approach of gestural phonology assumes a common gestural domain for both phonetic details and phonological structure, in which the constellations of language-specific gestural details *are* the phonological elements of the language.

The concerns about language-specific phonetic implementations are of central importance for understanding how the infant eventually comes to recognize the organizing principles of his or her native sound system. The infant initially has access only to the surface phonetic details of native speech and must ultimately come to perceive the relation between these details and the more abstract phonological structures of the ambient language. One might suppose that all possible phonological categories are innately given, but that only some are maintained by language-specific experience, the others dropping out for lack of environmental support. The difficulties with this notion are that phonological categories and contrasts are defined by their lin-

guistic functions in a given language, and that they show language-specific differences in their phonetic realizations. They *cannot* be innately given and then weeded out by experience, given the nonuniversal mapping between phonetic details and phonological structures. Rather, infants must start with the ability to detect a broad range of possible articulatory gestures in speech, from which they can discover the specific patterns by which the ambient language harnesses phonetic details to serve phonological functions.

The preceding discussion reflects the fact that although phonologists are concerned primarily with the speaker–hearer's phonological *competence*, their investigation of this capability has been founded nearly exclusively on examinations of *performance*; that is, how the presumed underlying phonological competence of the speaker yields observed phonetic patterns in speech. However, we are concerned here with perception; that is, how the listener moves in the opposite direction, from the phonetics of perceived speech to the phonological structure of a language. If phonetic implementations are language-specific rather than universal, then what exactly does the mature listener perceive in non-native phones and contrasts, and how does the infant discover native phonological structure, given the phonetic surface patterning of the ambient language?

Perceptual Learning and Perceptual Tuning to the Native Phonology

A perceptual system that has become attuned shows increased ability to pick up that information to which it has been sensitized (e.g., E. Gibson 1963, 1966, 1969, 1977, 1988, Gibson and Gibson 1972, Gibson and Gibson 1955). This involves both detecting critical distinctions that were previously unnoticed and passing over irrelevant differences that play no critical role toward the perceiver's goals. Perceptual learning entails discovering the *critically* distinctive features, the most telling differences among objects and events that are of importance to the perceiver. Information that does not serve this purpose tends not to be picked up.

Perceptual learning is discovering invariants in stimulation that reveal the structural and functional properties of the source objects and events. Often, higher-order invariants arise from lower-order invariants; there may even be several levels of lower-order 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 structure, may not be detectable until the perceiver has learned to pick up certain distinctive information at lower levels.

Educated speech perception should, therefore, actively seek and extract the invariants of language-specific articulatory gestures and

constellations of intergestural phasing at all levels from segments to syllables, words, and so forth. Language-specific gestural constellations are complex articulatory events, which are specified by higher-order invariants in the signal that automatically account for contextual variations such as speaking rate, speaker differences, and allophonic variation, so that perceivers "hear through" irrelevant lower-order variations. The converse of the efficiency of extracting higher-order invariants in native speech may be an increased difficulty in picking up lower-order gestural invariants of unfamiliar non-native categories that are irrelevant to critical native distinctions.

Language-specific invariants reflect the organization of distal articulatory events, but most or all of these higher-order invariants are initially beyond infants' perceptual reach. Infants still need to discover how simple articulatory gestures are harnessed by native speakers into higher-order gestural constellations. Infants actively explore utterances to discover the optimal gestural invariants that constitute the native language structures. The optimal invariants, of course, will change as the discovery of lower-order invariants permits the discovery of still higher-order ones. Developmental change in perception of native speech should be evident primarily as increasingly efficient pickup of critical differences. But perceptual learning also has implications for perception of non-native phonetic patterns. The latter changes are more dramatic because the attuned perceiver may be unable to find familiar gestural invariants in non-native speech.

A DIRECT REALIST MODEL OF CROSS-LANGUAGE SPEECH PERCEPTION

The remainder of the chapter describes the current state of a direct realist model of cross-language speech perception and development that I started developing several years ago. It should be noted that it is a working model, and, therefore, is subject to further revision. It rests on the premises of the ecological approach to speech perception, coupled with direct realist principles of perceptual learning. That is, it assumes that perceivers' attunement to native speech entails their discovery of the higher-order invariants that specify the gestural constellations of which the native phonological inventory is comprised. As perceivers become attuned to this language-specific information, they become increasingly adept and efficient at detecting the critically distinguishing properties of those constellations; that is, of the linguistically relevant contrasts among them. When attuned perceivers detect such higher-order invariants, they are picking up *reduced* information from stimulation; that is, the information they now gather from speech is more compact than was the full array of lower-order

simple gestural features that they initially detected before recognizing the patterns of coordination among them. It is especially this latter claim about perceptual attunement that is important to understanding how experience with the native language affects perception of non-native speech.

For heuristic purposes, the range of vocal tract gestures that are (or could be) harnessed by human languages, and their lawful perturbations of the energy transmitted through various media (acoustic, optic, tactile [pressure changes via elastic deformation of body tissue]), can be considered to define the contents of a "universal phonetic domain," akin to Catford's (1977) *anthropophonic space*. This "domain" refers to the spatial and temporal invariants of all possible phonetic gestures; that is, it is a multidimensional energy potential map of the dynamic formation of constrictions of varying degrees at various relative locations along the extent of the vocal tract. The universal phonetic domain is not, in this view, a mental construction resulting from the perceiver's extrapolation, reorganization, and interpretation of static dribs and drabs of proximal stimulus cues that are intrinsically meaningless, ambiguous, and inadequate. Nor is it the contents of a neural module wired with innate knowledge about the vocal tract. It refers, instead, to the types of gestures afforded by the physical structure and biodynamic constraints of the human vocal tract; it is assumed that perceivers recover information about actual speech gestures from the rich flow of lawfully shaped, multidimensional stimulation resulting from those gestures.

The naive perceiver detects, within the perturbation patterns of one or more energy media, information that specifies the simple gestural properties that occur in the speech of any language. This is because he or she has not yet discerned the more complex coordinations among such simple gestures that correspond to language-specific phonological elements such as phonemes, syllables and their constituents, and rhythmic units. Becoming attuned to such higher-order gestural constellations means increased efficiency at detecting in the energy flux the higher-order invariants that specify those constellations. Because higher-order invariants combine multiple lower-level details, they serve to reduce or optimize the information that perceivers must pick up by reducing the number of individual articulatory details that must be detected. Thus, once a highly attuned perceiver has discovered the more optimal higher-order invariants, the more numerous degrees of freedom describing the lower-order details will generally be passed over as being irrelevant in and of themselves with respect to his or her increasingly abstract goals. Attunement to the native language increases attention to the portions of the phonetic domain that the language has harnessed to serve linguistic functions; that is, detec-

tion of crucial information in these regions of the domain becomes both more sharply defined and more efficient. We can define this familiar region as "native phonological space." Metaphorically speaking, the gestural constellations that serve phonological functions in the native language become easily recognized as well-worn paths in the spatiotemporal terrain of the universal phonetic domain (see also, e.g., Lindblom 1992, Mohanon 1992).

Universal Phonetic Domain: Simple Gestures

The following brief conceptualization of the universal phonetic domain is based largely on the model of gestural phonology proposed by Catherine Browman and Louis Goldstein (1986, 1989, 1990a, b, c, 1992a, b). In some details, I have extended the description of the phonetic domain past the set of articulatory variables and parameter values handled by the computer model they have been developing to generate acoustic patterns from articulatory gesture specifications. I have done this in order to incorporate certain known gestural elements of languages other than English that are not yet incorporated into the computer model. In these cases, however, I have attempted to remain true to the spirit of gestural phonology.

The gestural model assumes that phonological patterning in languages obeys the constraints provided by the physical structure of the vocal tract and the movements that its biomechanical components afford. Therefore, the geometry upon which it bases its phonological descriptions is true to the vocal tract layout and its spatiotemporal movement possibilities. The basic claim is that the primitives, or "atoms," of phonological structure are articulatory gestures. Phonological structures are stable constellations, or "molecules," assembled with these atoms. A simple gesture is defined as the formation (and release) of some degree of relative constriction at some location along the vocal tract. Degree of constriction is either closed (e.g., stop consonants), critical (i.e., aerodynamic turbulence at the constriction: fricatives), narrow (glides, high vowels and low back vowels), mid (other vowels), or wide (velum for nasal stops, glottis for voiceless stops and fricatives). Location is jointly specified by vocal tract tube geometry and by articulatory geometry.

The tube geometry refers to the hierarchical spatial layout of the vocal tract as a series of connected tubes (see Figure 1). The root node is the vocal tract as a whole. At the next branching level, differentiated from the vocal tract node, are the glottal node and the supralaryngeal node. The state of the glottis makes the posterior end of the vocal tract tube either open or closed. Branching from the supralaryngeal node are the oral node and the nasal node. The nasal node refers to the large side branch of the supralaryngeal tube, which is open at the

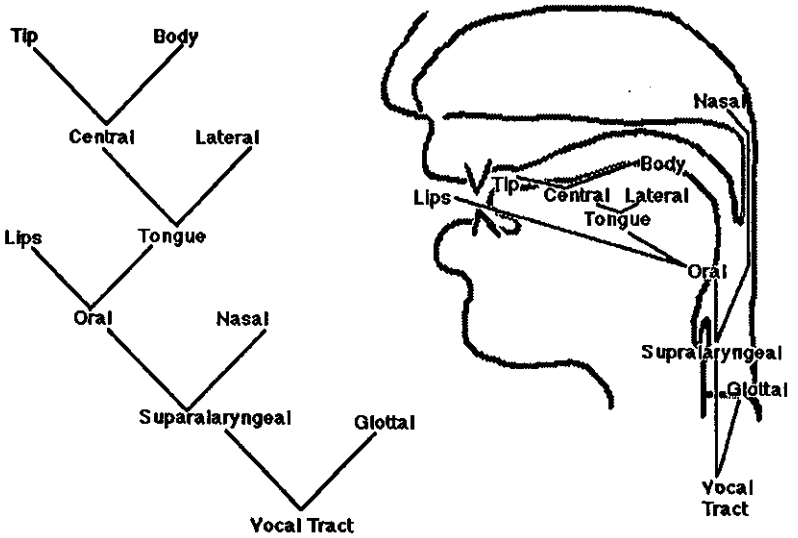

Figure 1. Tree diagram of the hierarchically organized tube geometry of the vocal tract. (Note that the tree diagram is reversed top-to-bottom, as compared to the typical downward-branching orientation of tree diagrams in linguistics, so that the relation between the tree structure and the geometry of the vocal tract can be seen better).

anterior end (nostrils) and may or may not be aerodynamically connected to the oral tube depending on whether its posterior end (velum) is opened or closed. The oral tube is further differentiated into the lips and the tongue. The lips may be used not only to produce varying degrees of constriction at the anterior end of the supralaryngeal tube, but also may be protruded to extend the length of that tube. The tongue node branches to the central and lateral nodes, which refer to differences in its positioning that can direct oral airflow along the center or the sides of the tongue, respectively. The central node is itself differentiated into a tongue-tip node and a tongue-body node, which can form constrictions of varying degrees at front-to-mid or mid-to-back locations, respectively, along the midsagittal extent of the oral tube.

The range of possible oral constriction locations is arrayed along the dorsal (upper) surface of the oral tube (from anterior to posterior): protruded lips, labial (nonprotruded), dental, alveolar, postalveolar, palatal, velar, uvular, and pharyngeal. These constrictions are formed by the movements of the active articulators, which are arrayed along the ventral (lower) surface of the oral tube. The glottal node and nasal node of the tube geometry have their active articulators as well: the vocal folds and the velum, respectively. The active articulators are

specified by the articulatory geometry. The root node, again, is the vocal tract. Subordinate to it are the glottal node, the velic node, and the oral node. At the glottal node, the vocal folds are the articulators responsible for the formation of varying degrees of constriction: fully closed for ejective or glottalized sounds, critical for voiced sounds, narrow for /h/, wide for aspirated voiceless sounds. In addition, the location of the glottis may be optionally changed from its normal resting position by raising the larynx (ejective stops) or lowering it (implosive stops). At the velic node, the velum is the active articulator and may be fully closed, wide open, or may show some degree of partial opening. Subordinate to the oral node are the lip and tongue nodes. There is further differentiation of the tongue node into tongue-tip and tongue-body nodes. For these two tongue nodes, tongue shape may be needed as an additional specification for some gestures (e.g., tongue grooving for /s/; side-to-side narrowing of the body for /l/; retroflexion of the tip for /ɹ/). Because some languages make use of more posterior constrictors than English does, it is necessary to assume a tongue-root node as a third subordinate to the tongue node, and although it is used quite rarely in the world's languages, an epiglottal node may also have to be assumed.

For oral node constrictions, the active articulators are arrayed along the lower surface of the supralaryngeal tube (anterior to posterior): lip, tongue tip, tongue body (dorsum), tongue root, and epiglottis. The lower lip may form constrictions against the upper lip (bilabial place of articulation) or the teeth (labiodental). The tongue tip may form constrictions at labial (linguolabial), dental (linguodental or interdental), alveolar, postalveolar (including some retroflex and grooved tongue shapes), or palatal locations (as in the laminal retroflex stop found in Tamil). The tongue body may form constrictions at the palatal, velar, uvular or pharyngeal places, whereas the tongue root and epiglottis may only effect pharyngeal places of constriction (see Figure 2).

Native Phonological Space: Selection of Simple Gestural Settings, Assembly of Gestural Constellations

One way in which languages might differ phonologically, and therefore in the native phonological spaces they circumscribe within the universal phonetic domain, could theoretically be in their selection of simple gestures. For example, although some languages such as Arabic employ uvular and pharyngeal locations for the primary places of consonantal constriction gestures, English does not. Therefore, we might suppose that simple uvular and pharyngeal gestures would be included within the boundary of native phonological space that is

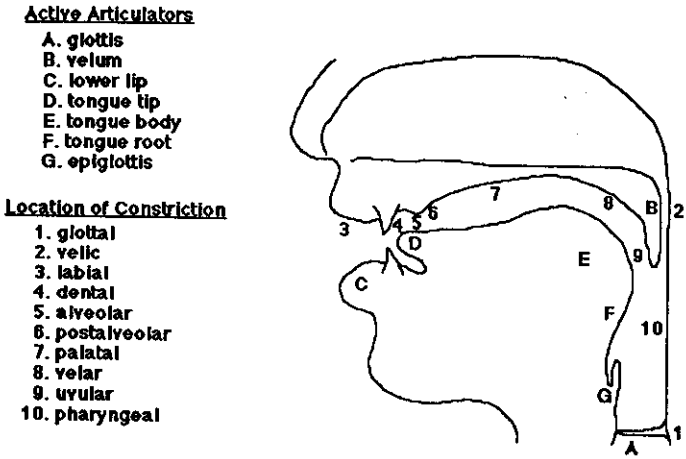

Figure 2. Schematic diagram of universal phonetic domain, with constriction locations noted numerically on the dorsal surface of the vocal tract tube and active articulators noted alphabetically on the ventral surface.

drawn by Arabic, but would not fall within the English boundary. The problem with this supposition, however, is that English *does* include uvular and pharyngeal constrictions as either secondary consonantal constrictions (e.g., uvular narrow constriction for /l/) or as primary vowel constrictions (pharyngeal narrow constriction for /a/). In other words, simple uvular and pharyngeal places of constriction must be incorporated into English phonological space.

In fact, a fundamental implication of the universal phonetic domain's spatial fidelity to the vocal tract is that the boundaries of a given native phonological space should automatically incorporate any "unused" places of articulation that fall between places that it does use. To put it simply: native phonological space doesn't gerrymander the upper surface of the oral tube. Thus, if pharyngeal and labial are employed as constriction places in the language, then uvular must also be within native phonological space even if the language employs no primary constrictions at the uvular place; likewise for dental place, and so on. That is, the interim positions will be perceived as possible locations for gestures that might serve as phonological units—they will seem speechlike, dependent, of course, on whether the associated degree of constriction also falls within native phonemic space (see next paragraph). Given that bilabial consonants and the low vowel /a/ are nearly universal in languages (e.g., Ladefoged and Maddieson 1990, Lindblom 1990, Lindblom, Krull and Stark 1993, Lindblom and Maddieson 1988, Maddieson 1984,

Stevens and Keyser 1989), the loci of constrictions along the upper surface of the oral tube should range from the lips to at least the upper pharynx in the native phonological spaces of the vast majority of languages. Conversely, most languages apparently draw the line for possible oral constriction locations at some point above the lower portion of pharynx, given the extreme rarity of epiglottal consonants in phonological inventories. As for the other nodes of the vocal tract tube, it would appear that almost all languages also include velic and glottal constriction locations in their native phonemic spaces, given the nearly universal appearance of at least one nasal stop and at least one voiceless (unaspirated) stop in the phonological inventories of languages.

Thus far we have only addressed how native phonological space may delineate a region in the universal phonetic domain along the static spatial dimension of constriction location, which corresponds to the enduring physical structure of the dorsal surface of the vocal tract. But this information alone is insufficient to define the bounds of native phonological space. We must also include the spatiotemporal definition of the simple constriction gestures themselves, which are local changes in tube diameter produced by the active articulators found on the ventral surface of the vocal tract. Degree of constriction may be viewed as a higher-order dynamic dimension in the universal domain that may be circumscribed differently by different languages. Nearly all languages produce constrictions with the lower lip, tongue tip, and tongue body. The tongue root may be actively engaged by fewer languages (although production studies are needed to ascertain this), and very few engage the epiglottis. This suggests that native phonological space for most languages will draw the line on the active-articulator dimension above the epiglottis, certainly, and perhaps above the tongue root.

As for simple gestural variations in constriction degree, virtually all languages employ fully closed constrictions (stops) as well as narrow constrictions (high vowels, low back vowels, glides) for oral gestures, closed and wide constrictions for velum gestures, and wide as well as critical constrictions for glottal gestures (the latter being the default mode for voicing). Of the simple unitary gestures, however, some languages fail to include critical constrictions (fricatives) or mid constrictions (nonhigh, nonlow vowels) for oral gestures. As was the case with constriction location, however, we should expect constriction degree not to be gerrymandered by the boundary of native phonemic space. Therefore, even if a language does not use critical oral constrictions, the critical constriction degree falls between the fully closed and narrow constriction degrees and so should be encompassed by any language that uses the latter two settings. That is, in virtually all lan-

guages, fricatives should seem speechlike, even if the language has no actual fricative phonemes.[2]

More important is the fact that native phonological space incorporates an even higher-order temporal dimension than those described so far for simple gestures. As stated earlier, Browman and Goldstein argue that phonological structures generally are not simple, single gestures but rather are stable assemblies of multiple gestures, which they refer to as gestural constellations. Languages differ most dramatically in their assembly of simple gestures into gestural constellations. Such phonological structures are certainly not constrained to just the segmental level of linguistic description, but extend to other phonological levels including syllables and their constituents, rhythmic units, etc. To an overwhelming extent, however, cross-language research on speech perception has focused on segmental contrasts and comparisons, so I likewise focus on segments and segmental contrasts in this discussion.

Two aspects of gestural constellations are important in the consideration of native phonological space. One is, of course, that languages are selective in the particular coordinations among simple gestures they employ to serve phonological functions. For example, French forms constellations between a narrow or mid constriction gesture of the tongue tip or body, and a wide velum gesture, to form nasalized vowels as phonological elements that contrast with nonnasalized vowels. However, although English uses some of the same tongue gestures for vowels, and uses velum gestures contrastively in constellations with other gestures, and also shows coarticulation of velum and tongue gestures when vowels are adjacent to nasal consonants, it does not form nasal vowel constellations to serve *phonologically contrastive* functions. The other important aspect of gestural constellations is that they generally depend upon specific phasing relations between gestures. For example, English assembles constellations that combine an oral closure with a glottal widening gesture at word-initial and stress-initial positions, forming aspirated voiceless stop consonants as phonological elements that contrast with voiced stops that lack the glottal opening gesture. For English voiceless aspirated stops, the glottal gesture begins at about 90° of phase angle into the oral closure gesture (i.e., the glottal gesture begins after the oral gesture does) and reaches its peak opening at the release of the stop closure. In

[2]Although this description of simple gestures has listed explicitly the locations of constriction and of active articulators, it should be noted that perceivers may actually recover higher-order invariants from the resonances in speech (formants), such as information about proportional lengths of the vocal tract on either side of the primary constriction. That is, rather than picking up low-level information about specific articulators and locations and constriction degrees, they may instead detect the relative position and aerodynamic influence of the constrictions within the vocal tube system.

French, however, voiceless stops are unaspirated, although they also couple a glottal opening gesture with an oral closure gesture. The language difference in this case rests primarily in the phasing of the gesture. In French, the glottal opening gesture begins simultaneous with the onset of the oral closure gesture, and the glottal gesture reaches its peak before the release of the oral closure. In addition, the French glottal gesture is smaller in magnitude than the English glottal gesture.

Thus, native phonological space for different languages is expected to reflect differences primarily in the patterning of gestural constellations that each language harnesses for phonological purposes. Languages differ in which gestures they combine and also in the particular phasings they set between the gestures used to form constellations.

Perceiving Non-Native Speech: Perceptual Assimilation Model (PAM)

Although the discussion thus far about the native phonological spaces of diverse languages has emphasized cross-language differences in the gestural constellations they assemble for phonological purposes, this should not leave the wrong impression that they have little in common gesturally. The fact is that, because they all draw upon the gestural possibilities of the human vocal tract (i.e., from the universal phonetic domain), there is usually a great amount of overlap among languages in the gestures and constellations contained within their individual phonological spaces, at least at the segmental level that is of interest to us here. As linguists have long noted, numerous segments (i.e., gestural constellations) are shared quite widely among languages, being identical or nearly so among them. Although these commonalities have understandably captured the theoretical interest of those concerned with universal properties of language and language ability, they do not particularly illuminate how attunement to one (or more) language(s) constrains the perception of non-native segments from unfamiliar languages. Non-native segments are those whose gestural elements or intergestural phasing do not match precisely any native constellations.

The fundamental premise of the perceptual assimilation model of cross-language speech perception is that non-native segments, nonetheless, tend to be perceived according to their similarities to, and discrepancies from, the native segmental constellations that are in closest proximity to them in native phonological space. Because the universal phonetic domain and native phonological space are defined by the spatial layout of the vocal tract and the dynamic characteristics of articulatory gestures, those distal properties provide the dimensions within which similarity is judged. For a native listener of a language that has no dental stop but does have bilabial, alveolar, and velar stops, the tongue tip constriction of the dental stop is straightforwardly closer

in native phonological space to the alveolar place than to the others, because the articulation involved is the same and the place of constriction is more similar than those of bilabial or velar stops. If the listener's language has no ejective stops but does have voiceless aspirated and prevoiced stops, the glottal gestures and phasings of ejectives are more similar gesturally to the voiceless aspirates than to the prevoiced stops, given that both glottal closure and glottal widening prevent voicing and that the glottal gestures of both are phased so as to reach their peaks at the release of the oral closures with which they are linked.

Similarity between non-native segments and native gestural constellations, as indexed by the spatial proximity of constriction locations and active articulators and by similarities in constriction degree and gestural phasing, are predicted to determine listeners' perceptual assimilation of the non-native phones to native categories. That is, the listener is expected to detect gestural similarities to native phonemes. At the same time, however, it is also expected that the listener will detect discrepancies from the gestural properties of native constellations as well, especially when the discrepancies are large. In very discrepant cases, the non-native sounds may be perceived only as having globally speechlike properties but may not assimilate strongly to any particular native category. In the extreme, they may not even be recognized as speechlike gestural constellations, but may instead be heard as nonspeech sounds, such as other types of sounds made by vocal tract events (e.g., choking) or as other humanly produced or even nonhuman events (e.g., fingers snapping or a cork popping, respectively). Assimilation is assumed to be tapped by tests that measure identification (labeling), classification, or categorization (including goodness ratings) of non-native phones. Specifically, the following patterns of perceptual assimilation of non-native segments have been outlined for the perceptual assimilation model (see also Best 1993, 1994, in press a):

1. *Assimilated to a native category:* clearly assimilated to a particular native segmental category, or perhaps to a cluster or string, in which case it may be heard either as:
   a. a good exemplar of that category
   b. an acceptable but not ideal exemplar of the category
   c. a notably deviant exemplar of the category
2. *Assimilated as uncategorizable speech sound:* assimilated within native phonological space as a speechlike gestural constellation, but not as a clear exemplar of any particular native category (i.e., it falls within native phonological space but in between specific native categories)

3. *Not assimilated to speech (nonspeech sound):* not assimilated into native phonological space at all; heard, instead, as some sort of nonspeech sound

Assimilation patterns for non-native *contrasts* follow predictably from the assimilation of each member of the contrast. Moreover, degree of perceptual differentiation, or discriminability, for diverse non-native contrasts is predictable from the assimilation of each of the contrasting non-native segments. A summary follows for most of the different pairwise assimilation patterns that are possible, and for the discrimination levels predicted for each:

Two-Category Assimilation (TC Type) Each non-native segment is assimilated to a different native category, and discrimination is expected to be excellent.

Category-Goodness Difference (CG Type) Both non-native sounds are assimilated to the same native category, but they differ in discrepancy from native "ideal" (e.g., one is acceptable, the other deviant). Discrimination is expected to be moderate to very good, depending on the magnitude of difference in category goodness for each of the non-native sounds.

Single-Category Assimilation (SC Type) Both non-native sounds are assimilated to the same native category, but are equally discrepant from the native "ideal"; that is, both are equally acceptable or both equally deviant. Discrimination is expected to be poor (although it may be somewhat above chance level).

Both Uncategorizable (UU Type) Both non-native sounds fall within phonetic space but outside of any particular native category, and can vary in their discriminability as uncategorizable speech sounds. Discrimination is expected to range from poor to very good, depending upon their proximity to each other and to native categories within native phonological space.

Uncategorized versus Categorized (UC Type) One non-native sound assimilated to a native category, the other falls in phonetic space, outside native categories. Discrimination is expected to be very good.

Nonassimilable (NA Type) Both non-native categories fall outside of speech domain being heard as nonspeech sounds, and the pair can vary in their discriminability as nonspeech sounds; discrimination is expected to be good to very good.

Empirical Support for Perceptual Assimilation Predictions

We have conducted several studies of non-native speech perception in adults to test the PAM, which are described in somewhat more detail

in several recent papers (Best 1993, 1994a, 1994b). The results have been quite supportive of the model's predictions, both with consonant contrasts and with vowels. Specifically, English-speaking adults were expected to perceive a variety of non-nasalized click consonant contrasts from Zulu as NA nonspeech sounds, because their double-stop articulations and suction release gestures are highly deviant from any English gestural constellations. As a result, English-speaking Americans' discrimination of these clicks was expected to be very good, despite the listeners' lack of prior exposure to click consonants in speech. Both the assimilation predictions and the discrimination predictions were upheld (Best, McRoberts, and Sithole 1988). In a second study, English-speaking adults were tested on three additional Zulu contrasts: a lateral fricative voicing distinction that was expected to assimilate to two different English categories (TC type), a velar voiceless aspirated versus ejective stop distinction that was expected to assimilate to a difference in goodness of fit to the English /k/ category (CG type), and a voiced bilabial plosive versus implosive distinction that was expected to assimilate as nearly equal exemplars of English /b/ (SC type, or weak CG type). All predictions were strongly upheld. The lateral fricatives showed TC assimilation and near-ceiling discrimination levels for almost all listeners, although most of them assimilated these sounds to clusters of English consonants such as "shl" and "zhl" rather than to single segments; many of the perceived clusters were nonpermissible in initial position in English. The velars showed clear CG assimilation and were easily discriminated but significantly less well than the TC contrast. The bilabials showed either SC assimilation or weak CG assimilation, and much lower discrimination than the other two contrasts (although performance was above chance). To test the generality of the TC performance pattern, the Ethiopian ejective contrast /p'/–/t'/ was presented to a second group of listeners. As predicted, the subjects were virtually unanimous in their assimilation of the Ethiopian phones to English /p/ and /t/, and yielded near-ceiling discrimination levels (Best 1990). Further support for the discrimination patterns predicted for differing non-native assimilation types has also been obtained in a study comparing American and Japanese listeners' categorical perception of three synthesized English glide contrasts (Best and Strange 1992).

More recently, we have extended the assimilation predictions to American listeners' perception of several non-native vowel contrasts. One of these, a Norwegian out-rounded versus unrounded contrast between high front vowels, yielded an SC assimilation pattern and poor discrimination. Two other contrasts (Norwegian high front unrounded versus in-rounded; French rounded high versus mid front vowels) were assimilated predominantly as TC contrasts and were

discriminated extremely well. Three other vowel contrasts (French rounded mid front versus central vowels; French oral/nasal back rounded vowels; Thai high versus mid back unrounded vowels) showed mixed assimilation patterns, varying from TC to CG to UC to SC types. The latter three contrasts were discriminated quite well also, but there was a strong association between individual listeners' actual perceptual assimilation patterns and their level of discrimination, so that discrimination was better for TC assimilations than for CG assimilations, which were, in turn, better than for SC assimilations (Best, Faber, and Levitt, in preparation).

We have also conducted a number of studies with English-learning infants to investigate the basis of developmental changes in perception of non-native contrasts. In contrast to Werker's reports that infants' early discrimination of many non-native contrasts gives way to lack of discrimination by 10–12 months (Werker et al. 1981, Werker and Lalonde 1988, Werker and Tees 1984; see also Werker and Pegg 1992, Werker and Polka 1993), we found that infants, like adults, continued to discriminate the Zulu click consonants up to at least 14 months of age (Best, McRoberts, and Sithole 1988). Also like adults, 10–12 month olds showed good discrimination of the Ethiopian ejective /p'/–/t'/ contrast, but poor discrimination of Zulu plosive versus implosive bilabials. However, they showed only marginal discrimination of the Zulu voiceless versus ejective velar contrast, and no discrimination of the Zulu lateral fricative voicing contrast, on which English-speaking adults had performed quite well (Best 1991, Best et al. 1990). The 10–12 month-olds' difficulty with the Zulu lateral fricatives, but not with the Ethiopian ejectives, was puzzling given that both contrasts had resulted in TC assimilation patterns by adults. In previous papers I suggested that perhaps their relative difficulty with the lateral fricatives was an indication that these older infants still were not perceiving segmental categories as phonological contrasts, but rather were focusing on the goodness of exemplars within native categories. In the case of the Ethiopian ejectives, the goodness of fit to English voiceless stops was likely to have been fairly good, as it was for adult listeners. However, the older infants may have failed to perceive the Zulu lateral fricatives as very good exemplars of any particular English categories—even the adults had shown a wide range of variation in exactly what naive transcriptions they gave these non-native sounds. However, a recent report by Jusczyk et al. (1993) suggests another possibility that is even more interesting from the perspective of phonological development. Those authors found that by 9 months of age, infants show consistent preferences for listening to lists of unfamiliar words that display native rather than non-native phonotactic patterns (see Juszcyk et al., this volume). Recall that English-speaking adults generally assimilated

the Zulu lateral fricatives to clusters rather than to single English segments, and that these clusters were often nonpermissible in English, especially in initial position. That is, they violated English phonotactic constraints. Perhaps the 10–12-month olds were simply reflecting their sensitivity to native phonotactic constraints when they failed to discriminate the lateral fricative voicing contrast that 6–8-month olds had been able to discriminate.

ADDITIONAL WORK ON THE PERCEPTUAL ASSIMILATION MODEL

Further work is needed on several aspects of the PAM. For one, its implications for learning-related changes in adults' perception of differing types of non-native contrasts as they acquire new languages (L2) containing those contrasts have yet to be worked out and tested. Note, however, that the model is quite amenable to experience-dependent adjustments in adults' perception of previously unfamiliar contrasts. The direct realist approach assumes that perceptual learning continues into adulthood. Nonetheless, it certainly expects the individual's history of experience to be reflected in the pattern and ease of learning new phonological patterns, particularly the native language-attuned listener's tendency to detect higher-order rather than lower-order gestural invariants in speech. Thus, systematic differences between adults' and young children's perceptual learning for previously unfamiliar segmental contrasts should be expected. Some of the questions that might be of interest to pursue include the following: Do multilinguals' languages share a common but expanded native phonological space, or separate (perhaps overlapping) phonological spaces for each language? Is L2 learning the result of phonetic exposure or linguistic usage? To what extent is L2 perceptual learning fostered by, or hindered by, other aspects of language use in the L2, such as the semantic, morphological, syntactic, and pragmatic characteristics of the language?

A second issue that needs further work is the development of a detailed, objective means for predicting assimilation patterns and discrimination of particular non-native contrasts, given the phonological properties of the listener's own language (including listener languages other than English). Recently, Carol Fowler suggested developing a gesturally based feature description of English and non-English consonants for the purpose of calculating similarity relations among the consonants in the matrix, in order to make mathematical predictions about more versus less likely perceptual assimilation patterns for non-native consonants. She pointed out that various current mathematical models of memory and recognition begin with the assumption that the formation and retrieval of remembered patterns depends upon the

degree of similarity between previously presented patterns and the current probe (e.g., Eich 1985; Hintzman 1986; Nosofsky 1986, 1988a, b; Nosofsky, Clark and Shin 1989). Although we reject the mental-representationalist assumptions of the theoretical underpinnings for the computational models, the similarity formulae they employ are actually theory-neutral and, of course, show substantial overlap. The predictive success of such models obviously depends upon the feature sets and feature values employed in the similarity calculations (whether or not they explicitly acknowledge this). For our needs, Fowler and I are modeling the features as closely as possible on gestural phonology, true to direct realist assumptions. No such attempt has been made to describe or model native versus non-native phonemic similarities on the basis of either pure acoustic features or the classic features of generative phonology (e.g., Chomsky and Halle 1968). We have begun this avenue of work, and recently collected similarity judgments on native and non-native consonants in CVs by American English-speaking listeners.

CONCLUSION

To summarize, I have described here the theoretical rationale for a perceptual assimilation model that makes a coherent set of predictions about how listeners will categorize, or assimilate, non-native phones with respect to the phonological categories of their native language, and how they will discriminate non-native contrasts. This model is founded on a direct realist approach to perception, in which articulatory gestures are assumed to be the perceptual primitives for speech perception, including the perception of non-native speech. It is further assumed that evidence of those gestures is available in the speech signal and is directly picked up by listeners, without need for recourse to innate knowledge of the vocal tract or to indirect cognitive processing of raw acoustic information. I suggested that the perceptual assimilation model can be extended to account for early developmental changes in perception of non-native contrasts, as well as for later perceptual changes that may occur as adults learn new languages. Further work is needed to develop and test the implications of the model for second language learning and to test its underlying assumptions about the gestural basis for assimilation and discrimination of non-native phones and contrasts. Work on some of these latter issues is currently in progress.

REFERENCES

Archangeli, D., and Pulleyblank, D. 1994. *The Content and Structure of Phonological Representations.* Cambridge, MA: MIT Press.

Aslin, R. N., Pisoni, D. B., and Jusczyk, P. W. 1983. Auditory development and speech perception in infancy. In *Infancy and the Biology of Development,* ed. M. M. Haith and J. J. Campos. New York: Wiley.

Best, C. T. 1984. Discovering messages in the medium: Speech and the prelinguistic infant. In *Advances in Pediatric Psychology.* Vol. 2, eds. H. E. Fitzgerald, B. Lester, and M. Yogman. New York: Plenum.

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

Best, C. T. 1991. 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, April.

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

Best, C. T. 1994a. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words,* ed. J. Goodman and H. C. Nusbaum. Cambridge MA: MIT Press.

Best, C. T. 1994b. Learning to perceive the sound pattern of English. In *Advances in Infancy Research,* ed. C. Rovee-Collier and L. Lipsitt. Hillsdale NJ: Ablex.

Best, C. T., Faber, A., and Levitt, A. in preparation. Association between adults' perceptual assimilation and discrimination of diverse non-native vowel contrasts.

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

Best, C. T., McRoberts, G. W., and Sithole, N. N. 1988. The phonological basis of perceptual loss for non-native contrasts: Maintenance of discrimination among Zulu clicks by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14:345–60.

Best, C. T., and Strange, W. 1992. Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20: 305–30.

Browman, C. P., and Goldstein, L. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3:219–52.

Browman, C. P., and Goldstein, L. 1989. Articulatory gestures as phonological units. *Phonology* 62:201–51.

Browman, C. P., and Goldstein, L. 1990a. Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics* 18:299–320.

Browman, C. P., and Goldstein, L. 1990b. Representation and reality: Physical systems and phonological structure. *Journal of Phonetics* 18:411–24.

Browman, C. P., and Goldstein, L. 1990c. Tiers in articulatory phonology, with some implications for casual speech. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, ed. J. Kingston, and M. E. Beckman. Cambridge, UK: Cambridge University Press.

Browman, C. P., and Goldstein, L. 1992a. Articulatory phonology: An overview. *Phonetica* 49:155–80.

Browman, C. P., and Goldstein, L. 1992b. Response to commentaries. *Phonetica* 49:222–34.

Catford, J. C. 1977. *Fundamental Problems in Phonetics*. Bloomington, Indiana: Indiana University Press.

Chomsky, N., and Halle, M. 1968. *The Sound Pattern of English*. New York: Harper and Row.

Clements, G. N. 1992. Phonological primes: Features or gestures? *Phonetica* 49:181–93.

Cohn, A. 1990. Phonetic and phonological rules of nasalization. *UCLA Working Papers* 76: May, 244.

Cohn, A. C. 1993. Nasalization in English: Phonology or phonetics? *Phonology* 10:43–81.

Diehl, R., and Kluender, K. 1989. On the objects of speech perception. *Ecological Psychology* 1:1–45.

Dooling, R. J., Best, C. T., and Brown, S. D. 1995. Discrimination of full-formant and sinewave /la-ra/ continua by budgerigars *(Melopsittacus undulatus)* and zebra finches *(Phoepila guttata)*. *Journal of the Acoustical Society of America* 97:1839–46.

Eich. J. M. 1985. Levels of processing, encoding specificity, elaboration, and CHARM. *Psychological Review* 92:1–38.

Fant, G. 1960. *Acoustical Theory of Speech Production*. The Hague: Mouton.

Fourakis, M., and Port, R. 1986. Stop epenthesis in English. *Journal of Phonetics* 14:197–221.

Fowler, C. A. 1986. An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics* 14:3–28.

Fowler, C. A. 1989. Real objects of speech perception: A commentary on Diehl and Kluender. *Ecological Psychology* 1:145–60.

Fowler C. A. 1991. Sound-producing sources as objects of perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America* 88:1236–49.

Fowler, C. A., Best, C. T., and McRoberts, G. W. 1990. Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception & Psychophysics* 48:559–70.

Fowler, C. A., and Dekle, D. J. 1991. Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 17:816–28.

Gibson, E. J. 1963. Perceptual learning. *Annual Review of Psychology* 14:29–56.

Gibson, E. J. 1966. Perceptual development and the reduction of uncertainty. *Proceedings of the 18th International Congress of Psychology* 7–17.

Gibson, E. J. 1969. *Principles of Perceptual Learning and Development*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Gibson, E. J. 1977. How perception really develops: A view from outside the system. In *Basic Processes in Reading: Perception and Comprehension*, eds. D. LaBerge and S. J. Samuels. Hillsdale, NJ: Erlbaum Associates.

Gibson, E. J. 1988. Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annual Review of Psychology* 39:1–49.

Gibson, E. J. 1991. *An Odyssey in Learning and Perception*. Cambridge, MA: Bradford Books (MIT Press).
Gibson, E. J., and Gibson, J. J. 1972. The senses as information-seeking systems. *(London) Times Literary Supplement* June 23:711–12.
Gibson, J. J. 1966. *The Senses Considered as Perceptual Systems*. Boston, MA: Houghton-Mifflin.
Gibson, J. J. 1979. *The Ecological Approach to Visual Perception*. Boston, MA: Houghton-Mifflin.
Gibson, J. J., and Gibson, E. J. 1955. Perceptual learning: Differentiation or enrichment? *Psychological Review* 62:32–41.
Gibson, E. J., and Gibson, J. J. 1972. The senses as information-seeking systems. *(London) Times Literary Supplement* June 23:711–12.
Goldstein, L., and Browman, C. P. 1986. Representation of voicing contrasts using articulatory gestures. *Journal of Phonetics* 14:339–42.
Gram, M. S. 1983. *Direct Realism: A Study of Perception*. The Hague: Martinus Nijhoff Publishers (Kluwer Academic Publishers Group).
Hintzman, D. L. 1986. "Schema abstraction" in a multiple-trace memory model. *Psychological Review* 93:411–28.
Jusczyk, P. W. 1993. From general to language-specific capacities: The WRAPSA Model of how speech perception develops. *Journal of Phonetics* 21:3–28.
Jusczyk, P. W. 1981. Infant speech perception: A critical appraisal. In *Perspectives in the Study of Speech*, ed. P. D. Eimas and J. A. Miller. Hillsdale, NJ: Lawrence Erlbaum Associates.
Jusczyk, P. W. 1986. Toward a model of the development of speech perception. In *Invariance and Variability in Speech Processes*, ed. J. S. Perkell, and D. H. Klatt. Hillsdale, NJ: Lawrence Erlbaum Associates.
Jusczyk, P. W. in press. Language acquisition: Speech sounds and the beginnings of phonology. In *Handbook of Perception and Cognition, Vol. 11: Speech, Language, and Communication*, ed. J. L. Miller and P. D. Eimas. Orlando, FL: Academic Press.
Jusczyk, P. W., Friederici, A. D., Wessels, J., Svenkerud, V. Y., and Jusczyk, A. M. 1993. Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language* 32:402–20.
Keating, P. A. 1984. Phonetic and phonological representation of stop consonant voicing. *Language* 60:286–319.
Keating, P. A. 1988. The phonology–phonetics interface. In *Linguistics: The Cambridge Survey. Vol. I: Grammatical Theory*, ed. F. Newmeyer. Cambridge, UK: Cambridge University Press.
Keating, P. A. 1990. Phonetic representations in a generative grammar. *Journal of Phonetics* 18:321–34.
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.
Ladefoged, P., and Maddieson, I. 1990. Vowels of the world's languages. *Journal of Phonetics* 18:93–122.
Liberman, A. L., and Mattingly, I. G. 1985. The motor theory of speech perception revised. *Cognition* 21:1–36.
Liberman, A. M., and Mattingly, I. G. 1989. A specialization for speech perception. *Science* 245:489–94.
Lindau, M. 1982. Phonetic differences in glottalic consonants. *UCLA Working Papers in Phonetics* 54:66–77.
Lindblom, B. 1990. On the notion of "possible speech sound." *Journal of Phonetics* 18:135–52.

Lindblom, B. 1992. Phonological units as adaptive emergents of lexical development. In *Phonological Development: Models, Research, Implications* ed. C. A. Ferguson, L. Menn, and C. Stoel-Gammon. Timonium, MD: York Press.

Lindblom, B., Krull, D., and Stark, J. 1993. Phonetic systems and phonological development. In *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, ed. B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, and J. Morton. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Lindblom, B., and Maddieson, I. 1988. Phonetic universals in consonant systems. In *Language, Speech, and Mind*, ed. L. M. Hyman and C. N. Li. London and New York: Routledge.

Löfqvist, A. 1980. Interarticulator programming in stop consonant production. *Journal of Phonetics* 8:475–90.

Löfqvist, A. and Yoshioka, H. 1984. Intrasegmental timing: Laryngeal-oral coordination in voiceless consonant production. *Speech Communication* 3:279–89.

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

Maddieson, I. 1984. *Patterns of Sound*. Cambridge: Cambridge University Press.

Mohanan, K. P. 1986. *The Theory of Lexical Phonology*. Boston: D. Reidel Publishing Company.

Mohanan, K. P. 1992. Emergence of complexity in phonological development. In *Phonological Development: Models, Research, Implications*, ed. C. A. Ferguson, L. Menn, and C. Stoel-Gammon. Timonium, MD: York Press.

Nosofsky, R. M. 1986. Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General* 113:39–57.

Nosofsky, R. M. 1988a. Similarity, frequency, and category representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14:54–65.

Nosofsky, R. M. 1988b. Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14:700–708.

Nosofsky, R. M., Clark, S. E., and Shin, H. J. 1989. Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15:282–304.

Pierrehumbert, J. B. 1990. Phonological and phonetic representation. *Journal of Phonetics* 18:375–94.

Pierrehumbert, J. B., and Pierrehumbert, R. T. 1990. On attributing grammars to dynamical systems. *Journal of Phonetics* 18:465–77.

Reed, E., and Jones, R. 1982. *Reasons for Realism: Selected Essays of James J. Gibson*. Hillsdale NJ: Lawrence Erlbaum Associates.

Stevens, K. N., and Keyser, S. J. 1989. Primary features and their enhancement in consonants. *Language* 65:81–106.

Studdert-Kennedy, M. 1985. Perceiving phonetic events. In *Persistence and Change*, ed. W. H. Warren, and R. E. Shaw. Hillsdale, NJ: Erlbaum.

Studdert-Kennedy, M. 1986. Development of the speech perceptuomotor system. In *Precursors of Early Speech*, ed. B. Lindblom, and R. Zetterstrom. New York: Stockton Press.

Studdert-Kennedy, M. 1989. The early development of phonological form. In *Neurobiology of Early Infant Behavior*, ed. C. von Euler, H. Forssberg and H. Lagercrantz. Basingstoke, England: MacMillan.

Studdert-Kennedy, M. 1991. Language development from an evolutionary perspective. In *Language Acquisition: Biological and Behavioral Determinants*,

ed. N. Krasnegor, D. Rumbaugh, R. Schiefelbusch and M. Studdert-Kennedy. Hillsdale, NJ: Erlbaum.

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.

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

Werker, J. F., Gilbert, J. H. V., Humphrey, K., and Tees, R. C. 1981. Developmental aspects of cross-language speech perception. *Child Development* 52:349–55.

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

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

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

Werker, J. F., and Polka, L. 1993. The ontogeny and developmental significance of langauge-specific phonetic perception. In *Developmental Neurocognition: Speech and Face Processing in the First Year of Life,* ed. B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, and J. Morton. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Zsiga, L. C. 1993. Features, gestures, and the temporal aspects of phonological organization. Unpublished Ph.D. dissertation, Yale University, New Haven, CT.