

**EMERGENCE OF LANGUAGE-SPECIFIC CONSTRAINTS IN
PERCEPTION OF NON-NATIVE SPEECH: A WINDOW ON EARLY
PHONOLOGICAL DEVELOPMENT**

CATHERINE T. BEST

*Department of Psychology
Wesleyan University
Middletown Connecticut
06459 U.S.A*

*and Haskins Laboratories
270 Crown Street
New Haven, Connecticut
06511 U.S.A*

ABSTRACT. Adults have difficulty discriminating many non-native speech contrasts, yet young infants discriminate both native and non-native contrasts. Language-specific constraints appear by 10–12 months. Evidence presented here suggests that mature listeners' discrimination is constrained by perceived similarities between non-native sounds and native categories, and that this native language influence may not be fully developed at 10–12 months. The findings suggest that young infants have broadly-tuned perception of phonetic details. Next, they begin to discern equivalence classes that roughly correspond to native phonemes. Perception of phonological *contrasts*, however, depends on recognition of their linguistic function, and thus develops later. But what sort of information in speech forms the basis for perception of equivalence classes or phonemic contrasts? I argue that distal articulatory gestures, rather than proximal auditory-acoustic cues or abstract phonetic features, are the primitives both for adults' perceptual assimilations of non-native phones and for infants' emerging recognition of native categories.

Introduction

Cross-language speech research has found that adults often have substantially greater difficulty discriminating unfamiliar non-native consonant contrasts than do the speakers of those other languages. Yet infants in the first few months discriminate a wide range of both native and non-native phonetic distinctions among consonants. Thus, the infant's speech discrimination ability is not initially influenced by whether the tested phonetic contrasts function as phonological oppositions in the ambient language. On the other hand, the mature listener's perception reflects constraints imposed, somehow, by the phonological system of the language s/he speaks. How do these language-specific constraints arise in development? How, exactly, do they begin to influence perception, and how might the constraints change developmentally? In other words, how might the emergence of language-specific constraints provide a window on the child's evolving knowledge about the phonological organization of the native language?

As Werker and colleagues have shown (1981, 1984, 1988, 1989) language-specific influences on discrimination of non-native consonant contrasts begin to appear sometime around 8 months of age, and are firmly established by 10–12 months. Her recent study with Polka (see Werker, this volume), and Kuhl's cross-language work on perceptual magnet effects (Kuhl et al., 1992), both indicate that language-specific influences on vowel perception emerge even earlier. I will return later to the potential importance of this and other differences between consonants and vowels. However, the balance of infant cross-language speech perception findings is tipped heavily toward consonants. My discussion will pertain primarily to consonant contrasts, but should also be relevant to other aspects of phonology.

Werker argued that the developmental change in perception of non-native contrasts is evidence *not* for a loss of sensory capacity (i.e., not a passive tuning of sensory mechanisms), but rather for a reorganization of phonetic perception (i.e., an active change in how phonetic information is handled). Specifically, she speculated, it reflects the infant's emergence from the initial pre-phonological processing of speech, and into the realm of phonology. I found that possibility intriguing, at least as much as the findings themselves. It started me puzzling over how infants might come to recognize *phonological* information. By 10 months, the average infant produces but a few, contextually-dependent words. And although 10-month-olds typically have at least a modest comprehension vocabulary, it apparently lacks minimal segmental contrasts, as does the productive lexicon for many months to come. Indeed, many current child phonology researchers have argued that broader articulatory patterns the size of syllables, words, or phrases, rather than the minimal-contrast features of standard segmental phonology, form the basis for early phonological behavior (e.g. Ferguson & Farwell, 1975; Macken, 1979; McCune & Vihman, 1987; Menn, 1983, 1986; Menyuk, Menn & Silber, 1986; Vihman & Velleman, 1989). Thus, if the perceptual reorganization in infants reflects the emergence of phonology, is that phonology a structured system of segmental contrasts, or does it have some other organization that may be displaced by yet further developmental change?

To approach this question, I began by posing another about the presumed developmental end-state: By what means does the native phonological system constrain the mature listener's perception of unfamiliar non-native contrasts? One possibility is that the listener's history of *differential phonetic exposure* during a critical developmental period could lead to maintenance or facilitation of physiological sensitivity to the acoustic properties of native segments (e.g., via tuning of neuronal feature detectors), and a complementary decline in sensitivity to those of non-native segments (e.g., Aslin & Pisoni, 1980). In this case the young infant, of course, would fail to show language-specific effects because it is still within the critical period and has had insufficient differential exposure. But there are several reasons why the sensory tuning explanation is unlikely to comprehensively account for language-specific effects on perception, as MacKain (1982) and I (Best, 1984, in press) have argued. To summarize briefly: First, mature listeners often *have* been exposed to segments or features which pose perceptual difficulties when they appear in non-native contrasts, but which appear in the listener's language as allophonic variants of native phoneme categories. That is, phonological function, rather than phonetic exposure, is what is critically lacking. Conversely, as I will describe later, certain other non-native contrasts pose little perceptual difficulty for adults, even when the native language *does not* offer allophonic exposure. Second, perception of at least some non-native phonemes can be improved by training or second language (L2) learning even in adulthood. Third, perception of non-native contrasts improves if task demands are reduced by various means. These observations are less compatible with critical period

sensory tuning than they are with more flexible, higher-level perceptual processes such as selective attention.

But if attention or some other higher process is tuned by the native language, what sort of information is it that shapes perception? Is it the physical, phonetic details of speech, its acoustic and/or articulatory patterns? Is it instead their linguistic function as phonological elements in the language? Or is it the *relation* between the phonetics and the phonology of the native language that shapes perception? Indeed, what is the relation between phonetics and phonology? This is not the sort of question typically addressed in infant speech perception research. Yet it is, I believe, at the core of how the native language influences perception of non-native phones, and of how a native phonology must emerge in development. The infant has access to the phonetic details of ambient speech, but what of its phonological structure? One might argue that the full set of humanly possible phonological categories is specified innately, and need only be weeded out by language-specific experience. But the problem is that phonological categories and contrasts are defined by their linguistic functions in specific languages, and differ in their phonetic realizations between languages -- indeed, the realizations differ among dialects, and undergo historical changes, even *within* languages. Thus, they *cannot* be innately given, any more than the universe of all possible words or morphological forms could be innately given. Rather, the infant must come equipped with the means by which to discover how the ambient language harnesses phonetic details to serve linguistic purposes (see also Jusczyk & Bertoncini, 1988). That is, infants must start with broad sensitivity to detect a range of possible phonetic patterns in speech, and on that foundation must come to discern the specific patterns selected by the native language to serve phonological functions. Yet how do infants get from phonetics to phonology? That depends on the relation between the two levels of information.

The standard linguistic assumption has been that phonology and phonetics are distinct domains; this is so even in nonlinear phonologies (e.g., Cohn, 1990). By that view, phonology refers to the language-specific patterning of discrete, timeless (hence abstract and static), qualitative linguistic representations, whereas phonetics refers to the quantitative and dynamic physical properties of speech acoustics, production, and perception, as realized in time and space. The mapping between linguistic, phonological contrasts and physical, phonetic events has been a central issue in phonology, and is in no sense trivial given the basic dimensional differences between the two domains as defined. Chomsky and Halle's (1968) widely-accepted proposal in *The Sound Pattern of English* was that phonetic implementation of phonological units is universal and non-linguistic. Coarticulation -- the overlapping or blending of neighboring phonological segments in actual speech -- and other aspects of the phonetic realization of abstract phonological elements were believed to follow from general physiological and/or mechanical principles, and so should occur cross-linguistically. Thus, phonology involves language-specific rules and is part of the linguistic grammar, whereas phonetic implementation was assumed to be automatic, universal, and non-linguistic. But more recent evidence has shown that phonetic implementations may be language- (and dialect-) specific as well, not purely mechanical (see, e.g., Fourakis & Port, 1986; Mohanan, 1986). For example, in Navajo ejective /p/ the delay and hence the "forcefulness" of the glottal release are greater than in Quechua /p/ (Lindau, 1982). Languages also differ in whether, and how much, vowels preceding voiced stops are lengthened relative to those before voiceless stops (Keating, 1984). Even related dialects display phonetic differences for the same phoneme, as in the delayed velum-lowering of nasalized vowels in Canadian French relative to European French (van Reenen, 1982).

The latter observations pose difficulties for automatic, universal mapping from phonological segments onto the phonetic output, the task with which phonologists have primarily been occupied. But my concern here is with perception, which presumably entails mapping from phonetics to phonology. If the phonetic implementations are language-specific rather than universal, then what information does the mature listener perceive in non-native phones and contrasts? More important, how could the infant discover the native phonology, given the phonetic input? In their discussions of problems with the standard phonological assumptions, both Fowler (e.g., 1980, 1986, 1989), and Browman and Goldstein (1986, 1989), have proposed that the primitives of both the phonetic level and the phonological level fall within a *single* domain of speech gestures, rather than in separate domains. Specifically, they propose that the phonological as well as the phonetic elements are comprised of the dynamic articulatory patterns in speech. Browman and Goldstein's gestural phonology is designed to capture (language-specific) coordinations among articulatory gestures such as lip closure and vocal fold vibration, including the language-specific timing or phasing relations among individual gestures. The phonological units thus have *intrinsic* temporal and spatial characteristics, rather than being abstract, timeless representations. Moreover, they are not constrained to the size of phonemic segments but may be larger (or perhaps even smaller).

Because it places the phonology of a language, which reflects the linguistic functions of its sound system, in the same concrete and dynamic domain as the phonetic substance of its spoken form, gestural phonology is an appealing vantage point from which to consider cross-language influences on perception, as well as how the infant may get from phonetics to phonology in the native language. A common domain, such as articulatory gestural information, would serve as a simple, direct link between perception and production of speech, which is central to both speech imitation and language acquisition. Specifically, I propose that both infant and adult listeners detect evidence in speech about the articulatory gestures of the vocal tract that produced the signal, consistent with Fowler's arguments that perceivers recover information from speech (and other sound-producing events) about the distal object and actions that produced the sounds (e.g., Fowler, 1986, 1989, 1991; Fowler, Best & McRoberts, 1990). It is important to note that by this, I mean that the articulatory gestures of speech are *directly perceived* and not that they are inferred from, or cognitively imputed to, the superficial acoustic properties of the signal. Because the speech signal is molded by the shape and movements of the vocal tract, according to the laws of physical acoustics, evidence about those properties is necessarily present in the patterning of the speech signal. And as Fowler and I argue, that evidence is available for perceivers to detect, as structured information about the distal vocal tract and events that produced the signal. Note also that this view does *not* make the motor theory assumption (e.g., Liberman & Mattingly, 1985) that listeners perceive speech by reference to the motor control, or the acoustic output, of their own speech. The claim here is only that listeners perceive information regarding the properties of articulatory gestures as produced by the speaker, whether or not the listeners could themselves produce similar signals (see also Fowler et al., 1990).

The notion that listeners perceive gestural information in speech is compatible with cross-modal speech perception findings. For example, McGurk (McGurk & MacDonald, 1976) and others have shown that when discrepant consonants are displayed in synchronized audio and dynamic video presentations of CV utterances, listeners do not detect the specific disparate information from the two modalities, but rather perceive a unified phonetic pattern that is compatible with the phonetic information from both modalities. This finding suggests that the two modalities convey

information about a common, underlying dimension such as articulatory gestural patterns. The alternative proposition that the dimensions may have become perceptually linked through learned associations is refuted by two recent reports. Fowler & Dekle (1991) showed that the cross-modal integration described by McGurk and MacDonald does occur for discrepant consonants in synchronized audio and dynamic tactile presentations of CVs (i.e., the Tadoma method of using the hand to feel movements of the speaker's articulations), even though normal listeners have had virtually no prior tactile-auditory experience with speech. However, the cross-modal effect disappears when discrepant consonants are synchronously presented in audio and written CVs, although literate subjects have had extensive associative experience with heard speech and its graphemic representation. In another recent study, Walton & Bower (in press) presented English-learning infants with audio repetitions of French /y/, a lip-rounded front vowel that does not occur in English, synchronous with dual silent videos of the English lip-rounded back vowel /u/ and the unrounded front vowel /i/. The infants' fixation patterns indicate that they matched /y/ with the lip-rounding pattern of /u/ even though they had not previously been exposed to /y/ with its visible pattern of lip-rounding.

I relate the gestural phonology approach to adults' perception of non-native contrasts in the proposal that mature listeners' percepts are guided by gestural similarities and dissimilarities between non-native phones and their own native phoneme categories. To illustrate this approach, I will rely on Browman and Goldstein's (e.g., 1989) way of schematically representing the gestural organization of an utterance. They start from articulatory data recorded during actual speech productions (i.e., changes in lip, tongue, and velum position over time), and interpret these data as pertaining to the gestural goals of forming and releasing constrictions of varying degrees and at various positions along the vocal tract. They then derive a *gestural score* which schematically represents the phasing, or relative timing, of coordinated articulatory gestures, of specific constriction degrees and at specific places, as these constrictions are formed by the diverse articulators of the vocal tract (e.g., lips, tongue tip, tongue body). I have taken their gestural score approach as a model for representing gestural properties of native and non-native phones, but have modified it to include a tongue-sides tier for specifying lateral gestures, and a larynx tier rather than a glottal tier to permit specification of laryngeal position as well as vocal fold state.

Figure 1 illustrates the schematic gestural score of English /ta/ (phonetically, [tʰa]) in the lower left panel. In it, a complete closure is executed by the tongue tip at the alveolar ridge, approximately coincident with the onset of a glottal widening gesture (abduction of the vocal folds) at the larynx. The glottal gesture is held past the release of the alveolar constriction and into the early portion of the pharyngeal narrowing gesture (tongue body/root) that is associated with the low back vowel /a/; this results in voiceless aspiration during the transition from the /t/ release and into the /a/. As the lower right panel of Figure 1 shows, the critical gestural difference between English [tʰa] and [pʰa] is that in the latter syllable the consonant constriction is bilabial, involving the two lips; the other two gestures, and the intergestural phasing, are essentially identical otherwise between the two syllables.

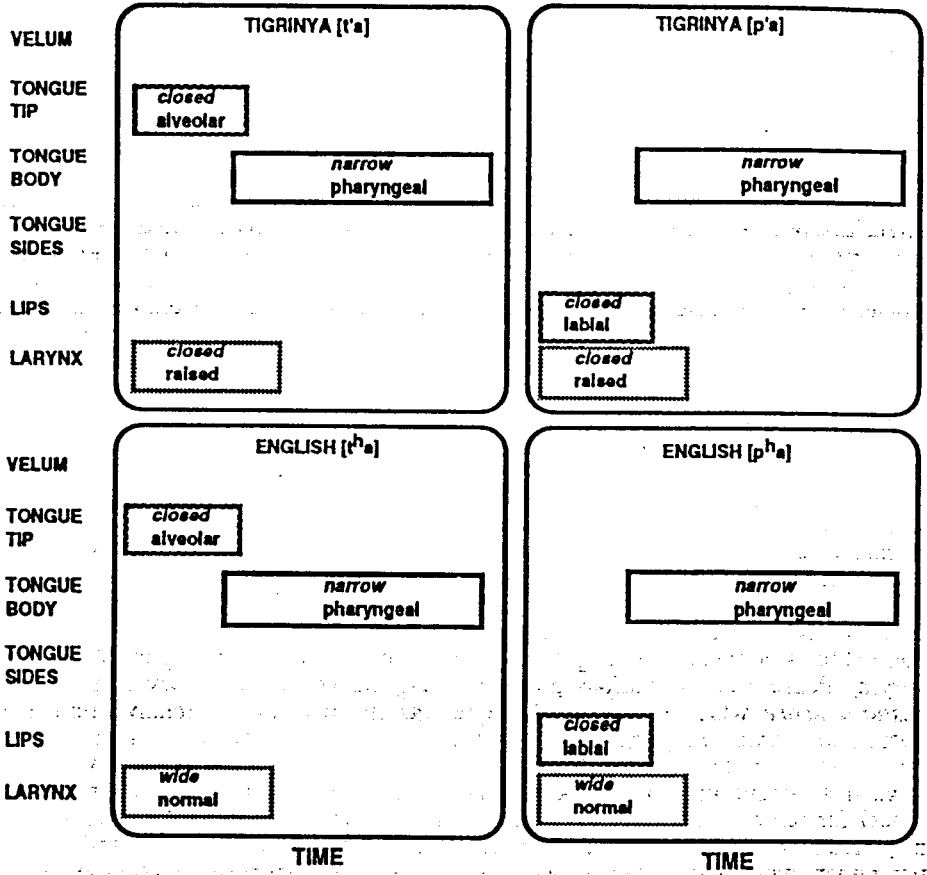


Figure 1. Schematic gestural scores for Ethiopian Tigrinya [t'a] and [p'a] (top), and for English [t^ha] and [p^ha] (bottom). Solid outlines around individual gestures indicate gestures that are identical in both contrasting phones and both languages. Dark dashed outlines indicate gestures that differentiate the contrasting phones in each language, but do not differ between languages. The gray stippled outlines indicate gestures that are constant within each language, but differ between languages.

To relate non-native phones to the native phonology, let us first consider the general layout of the native phonetic space. Table 1 summarizes the perception of non-native phones with respect to native phonetic space. I characterize native phonetic space as the realm of gestures that are employed by the native language, or that are reasonably similar to those gestures. Within this space are "islands" of gestural coordination patterns that serve linguistically as units of minimal contrast in the native phonology. There is assumed to be a range of within-category variation around the most common or typical coordination pattern(s) of each gestural-coordination island,

reflecting some phonetic differentiation in productions for each category, and in the corresponding perceptual structure for the category (see also Kuhl et al., 1992; Miller & Volaitis, 1989). The gestural properties of some phones from non-native languages will provide a phonetic match to native phonemes; these should count as native phones. Other non-native phones will fail to match exactly, yet will nonetheless be relatively similar, to the gestural coordination patterns of particular native categories. These will fall either within range of "good" (acceptable) tokens, or else somewhere more peripheral in that category. In the former case, the native category may exert a perceptual magnet effect (Kuhl et al., 1992), making the non-native sounds (nearly) indistinguishable from native ones (see also Flege, 1990). Non-native phones that are nearer the periphery of the native category, however, will be heard as relatively deviant tokens of that category. In the latter cases, variations in degree of discrepancy from native tokens should be perceptible. It is also assumed that some, perhaps many, of the possible gestural coordinations within general phonetic space will fall outside of any established native categories (see Flege, 1988). Such non-native phones involve individual gestures that are similar or identical to gestures produced in the native language, yet as phones they fail to correspond even moderately to any of the specific patterns of gestural *coordination* over time (i.e., *phasing* among several gestures) that are used linguistically in that language. These should fall in the uncommitted space between native categories, but nonetheless within general native phonetic space. That is, they will be perceived as "speech-like," but will not be classifiable as specific native phones. Still other non-native phones, however, may display gestural properties that are more obviously deviant from any gestures used in the native phonology. These should fall outside of native phonetic space altogether, and be heard as non-speech sounds. Non-phonetic space also includes coughs, vegetative sounds (e.g., eating, swallowing, breathing, etc.), and other vocalizations (e.g., imitations of animal sounds and of sound-making objects) that are too discrepant from gestures utilized by the native phonology to be perceived as speech-like.

Table 1. Perception of Non-Native Phones re: Native Phonetic Space

- 1. within native phoneme category**
 - a. phonetic match: native exemplar
 - b. phonetically similar: good exemplar
 - c. phonetically discrepant: deviant exemplar
- 2. in uncommitted phonetic space**
- 3. in non-phonetic space**

The perception of individual non-native phones according to the gestural properties of the native phonology should lead, in turn, to predictable differences in how non-native *contrasts* may be assimilated to native phonological contrasts, which should result in differences in discriminability for diverse non-native contrasts (see Table 2). If the two contrasting non-native phones are each perceived as an exemplar of a different native category, then discrimination should be quite high. The top panels of Figure 1 illustrate such a Two-Category (TC) assimilation pattern, in which the ejective contrast /t'-/p'/, from Ethiopian Tigrinya, is expected to assimilate to the English voiceless stop contrast /t-/p/ (bottom panels) on the basis of similarities in gestural coordination and in the critical gestural contrast. If, instead, both members of the non-native contrast are perceived as exemplars of the same native category, but

differ in Category Goodness (CG) within that native category (e.g., one is a "good" exemplar and other is deviant), then discrimination will be good but lower than for a Two-Category assimilation. In fact, discrimination of Category-Goodness differences should vary with differences between the two non-native phones in the magnitude of their discrepancy from native tokens -- discrimination will be relatively low if there is only a small difference in discrepancy between the non-native phones. At the extreme, both non-native phones could be perceived as equally deviant (or equally good) exemplars of a Single Category (SC) in the native phonology, and should thereby be quite difficult to discriminate. If the contrasting non-native phones both fall in uncommitted phonetic space, they will be Uncategorizable (UNC), and discrimination will likely be modest to poor. That is, discriminability will vary with degree of discrepancy between the two non-native phones in uncommitted phonetic space, but should be biased toward low performance. But if the contrasting non-native phones fall outside of native phonetic space altogether, they will be Non-Assimilable to the native phonology and will be heard as nonspeech sounds; in this case discrimination should be good but should vary dependent on the magnitude of auditory differences between the phones. Four of these assimilation patterns for non-native contrasts have been described elsewhere (Best, in press; Best & Strange, 1992); however, the characterization of uncommitted phonetic space, and of Uncategorizable contrasts, have not been described before. I should also note that non-native phone pairs can cut across assimilation types, although this possibility has not been addressed before, and was omitted from Table 1 for simplicity. For example, one phone may fall within a native category whereas the contrasting phone may fall in uncommitted phonetic space or in non-phonetic space. In such cross-category cases, discrimination should be quite good.

Table 2. Assimilation Effects on Discrimination of Non-Native Contrasts

<u>Contrast Assimilation</u>	<u>Discrimination Effect</u>
Two-Category (TC)	excellent discrimination each non-native phone assimilated to a <u>different</u> native phoneme category
Category-Goodness (CG)	good to moderate discrimination both non-native phones assimilated to the <u>same</u> native category, but differ in discrepancy from native phone
Single-Category (SC)	poor discrimination both non-native phones assimilated to the <u>same</u> native category, but are equally distant from native phone
Uncategorizable (UNC)	poor to moderate discrimination both non-native phones fall within uncommitted phonetic space
Non-Assimilable (NA)	good to moderate discrimination both non-native phones fall outside the bounds of native phonetic space and are heard as non-speech

Many of the contrastive perceptual assimilation predictions have been supported in studies conducted in my laboratory with English-speaking adults, as well as in a study with Japanese-speaking adults. As predicted, Americans perceived click consonants from Zulu as Non-Assimilable (NA) nonspeech sounds, and discriminated minimal click contrasts fairly well even though spoken English provides no exposure to clicks (Best, McRoberts & Sithole, 1988). Another study involved three additional Zulu contrasts (Best, 1990). One was a voicing distinction between lateral fricatives /ɬ-/ɮ/, in which tongue position is similar to English /l/ but manner of articulation corresponds more to the English /s-/z/ distinction. This Zulu contrast was expected to be assimilated to Two Categories (TC) in English, and thus to be discriminated quite well. A second Zulu contrast, that between the voiceless velar stop /k/ and the ejective velar stop /kʰ/, was expected to be assimilated as a Category Goodness (CG) difference in the English /k/ category. The third Zulu contrast was between the voiced plosive bilabial stop /b/ and its voiced implosive cognate /ɓ/, which were expected to be assimilated as showing little or no difference within the Single Category (SC) of English /b/. The results upheld the predicted order of performance in discrimination TC > CG >> SC. We also replicated the pattern of very high discrimination levels for another TC contrast from Ethiopian Tigrinya -- the ejectives /pʰ/ and /tʰ/ shown at the top of Figure 1, which were assimilated to English /p/ and /t/.

Likewise, we found that Japanese listeners' performance on synthetic stimulus series for three English glide consonant contrasts was consistent with assimilation predictions based on phonetic gestural similarities and discrepancies re: Japanese phonological categories and contrasts (Best & Strange, 1992). /w-/j/ is a phonemic distinction in Japanese and in English, but Japanese /w/ shows little of the lip-rounding seen in English, and tongue height for /j/ also differs (i.e., a TC contrast). The /w-/r/ distinction is phonemic in English but not in Japanese, where /w/ is similar to Japanese /w/ but English /r/ is dissimilar to Japanese /w/ (or /r/) (CG contrast). Both members of English /l-/r/ are quite discrepant from Japanese /r/ (SC or UNC contrast). As predicted, Japanese-speaking listeners labeled /w-/j/ quite categorically, like Americans, but placed their category boundary differently, at a position appropriate to the fact that Japanese /w/ is unrounded. The Japanese labeled /w-/r/ less consistently than Americans, and /r-/l/ least consistently. Their discrimination of the three contrasts followed the same order of performance levels: /w-j/ > /w-r/ > /r-l/. Interestingly, the subgroup of Japanese who had had intensive English conversation training or practice showed labeling and discrimination patterns that were more similar to the Americans (though not identical) than were those of the Japanese subgroup who had had little or no English conversation experience. In all, then, our findings with adults from two language communities strongly support the perceptual assimilation model.

Let us return now to the developmental shift in infants' perception of non-native contrasts at 10-12 months of age. How does that perceptual change relate to the adult performance pattern across varying non-native contrasts, and what can a comparison of older infants and adults from the same language environment tell us about the beginning development of a phonological system? Several alternative predictions may be offered about the underlying principle of reorganization in the 10-12 month olds' response to non-native consonant contrasts, as summarized in Table 3. Perhaps the developmental shift is motivated by some general cognitive principle, such as a **familiarity effect** in which sounds to which the infant has not been exposed become difficult to discriminate (but see the earlier argument against the differential exposure argument). In that case, we would expect all non-native contrasts to become difficult for the older infant, except perhaps for the Category Goodness (CG) type of

assimilation, where one phone is like a familiar native category while the other is clearly deviant and unfamiliar. On the other hand, perhaps the developmental shift reflects the infant's entry into using the mature organization of the native **phonological system**, in which case the older infant should show the same discrimination pattern across varying non-native contrasts as adults do: TC > CG = NA >> SC = UNC. Alternatively, phonological organization may have begun by 10–12 months yet still be immature in one of two ways. One such possibility is that infants' perception at this point does show reorganization around **phonemic contrasts**, but that the internal phonetic structure of phoneme categories is less differentiated than in adults. In this case, we should expect that TC (and NA) assimilation types would be discriminated well, but that CG, SC and UNC contrasts should become difficult for 10–12 month olds. Finally, immature phonological organization at that age may reflect **category recognition**, or perception of within-category phonetic structure, but may not yet show clear *linguistic* organization around minimal phonological contrasts. In the latter case, SC and UNC assimilations should become difficult for 10–12 month olds, whereas NA contrasts and CG contrasts between "good" and deviant exemplars of a native category should pose no difficulty. Some, perhaps many, TC contrasts should also remain discriminable, *if* the infant perceives the fit between the properties of the non-native phones and the structure of native categories. However, differences between native categories would not be perceived by the infant as functional, phonological contrasts. Some TC contrasts may become difficult for older infants to discriminate, if they fail to detect correspondences between the non-native phones and the phonetic structure of native categories.

Table 3. Hypotheses about the 10–12 Month Shift in Perception of Non-Native Contrasts

Familiarity Hypothesis

unfamiliar phones no longer command sufficient infant attention for discrimination

Predictions: poor discrimination for all non-native contrast types, alternatively,
good discrimination only for CG differences

Phonological System Hypothesis

mature phonological system like adults

Predictions: excellent discrimination for TC contrasts
good discrimination for CG and NA contrasts
poor discrimination for SC and UNC contrasts

Phonemic Contrast Hypothesis

perception of phonological *contrast*, but category details are relatively undifferentiated

Predictions: good discrimination for TC and NA contrasts
poor discrimination for SC, CG, and UNC contrasts

Category Recognition Hypothesis

perceptual focus on details of individual categories, not on phonological *contrasts*

Predictions: good discrimination for CG and NA contrasts
poor discrimination for SC and UNC contrasts
good discrimination for some TC contrasts but poor for others

To date, we have tested English-learning infants' discrimination for most of the non-native contrasts used in the adult studies, using a conditioned-fixation habituation procedure. In the first study, infants from 6 to 14 months continued to discriminate a Zulu click contrast that adults had heard as a NA contrast and discriminated relatively well (Best et al., 1988). This maintenance of click discrimination past 10–12 months is inconsistent with the predictions of the familiarity hypothesis. Next, we replicated this pattern of click discrimination in 6–8 and 10–12 month olds, as well as Werker's earlier finding (1984) of a decline between 6–8 and 10–12 months in discrimination of the Salish (northwest Native American) ejective velar–uvular stop contrast /k'–q'/, which constitutes a SC assimilation type for adults. We have since tested 6–8 and 10–12 month olds on the other three Zulu contrasts described earlier in our studies with adults: the TC lateral fricatives, the CG voiceless vs. ejective stop contrast, and the SC plosive vs. implosive bilabial stop contrast. Both infant ages discriminated the CG voiceless–ejective /k'–k'/. The older infants failed to discriminate the SC plosive–implosive /b'–b/, on which the younger group showed only marginal discrimination. More important, however, both ages failed to discriminate the contrast on which adults had shown their best performance, the TC lateral fricative /ɬ'–ɮ'/, which was particularly difficult for the 10–12 month olds. Interestingly, this difficulty persists even at 4 years, in contrast with good discrimination for /k'–k'/ (Insabella & Best, 1990). In a recent infant follow-up using a somewhat more stringent habituation criterion, 6–8 month olds *did* discriminate the TC lateral fricatives that were so easy for English-speaking adults, but the 10–12 month olds still failed utterly. However, the 10–12 month olds' difficulty apparently does not extend across all non-native contrasts that adults assimilate as TC contrasts, because they did discriminate the TC Tigrinya /t'–p'/ contrast described earlier (Best, 1991).

In summary, the infant studies show some similarities, but also some striking differences, in the non-native discrimination patterns of 10–12 month olds as compared to adults. There is clear evidence of some sort of language-specific attunement in speech perception by 10–12 months, but perception of non-native consonant contrasts has still not taken adult form. Thus, the infants' perceptual shift appears most consistent with the category recognition hypothesis summarized earlier. That is, older infants have begun to discern the phonetic properties of individual native categories, and this constrains perception of non-native contrasts. They recognize in some non-native phones, but not in others, certain coordinated phonetic patterns that they have begun to appreciate in native phones. However, they do not yet recognize functional, linguistic contrasts between segments -- a fundamental organizing principle of mature phonological systems.

Conclusion

In conclusion, I tentatively offer a sequence of phases in the development of the native phonology, based on the view from the window of age-related shifts in infants' perception of non-native contrasts (see Table 4). During the first phase, infants detect information in the speech signal regarding simple articulatory gestures produced by the speaker. This language-universal tendency gives way to language-specific effects in the second phase, when infants begin to recognize patterns of gestural *coordination* recurring in native speech, which may roughly correspond to phoneme categories. But during this phase infants do not yet perceive these patterns as participating in functional linguistic contrasts. The transition to language-specific constraints on perception of within-category phonetic structure is evident by 10–12 months for

consonants but may appear by 6 months for vowels (see Kuhl et al., 1992; Werker, this volume); possible explanations for this class difference are discussed below. In the third phase, recognition of functional linguistic contrasts may emerge around 18–24 months, as infants begin to notice and exploit the contrastive principles of their language, including its syntax and morphology. But awareness of phonemes as discrete, recombinable and commutable units may not appear until even later, around 5–6 years. Phonemic awareness appears to be helpful in acquisition of reading skill (e.g., Brady, Shankweiler & Mann, 1983; Liberman, Shankweiler & Liberman, 1989), and/or may itself be fostered by reading acquisition.

Table 4. Developmental phases in native phonological influences on speech perception

- | | |
|-----------------------------------|---|
| 1. (0 to 6–10 months): | extraction of simple articulatory gestural information
<i>language-universal phonetic details</i> |
| 2. (6–10 months to 18–24 months): | recognition of native patterns of gestural coordination
<i>language-specific phone categories</i> |
| 3. (18–24 months to 5–6 years): | emergence of functional phonological contrasts
<i>language-specific linguistic contrasts</i> |
| 4. (5–6 years to puberty [?]): | development of phonemic awareness
<i>language-specific phonemic units</i> |

Note the possible developmental differences described for vowels versus consonants, which I mentioned briefly at the outset of this chapter. This developmental difference for the two phonetic classes, if verifiable, has potential importance for understanding the development of native phonological influences on phonetic perception. It might reflect differences in categorical perception and/or short term memory for vowels versus consonants, which have long been noted in adults (e.g., Crowder, 1971; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Because performance on vowels can be shifted toward that found with consonants when the vowels are masked (e.g., by noise) or shortened in duration, it has been argued that the differences in categorical perception are attributable to differential decay of auditory memory for the transient acoustic properties of consonants versus the more sustained properties of vowels (e.g., Fujisaki & Kawashima, 1970; Pisoni, 1973; Stevens, 1968). Thus, it is possible that this memorial effect alone may account for the infant's earlier attunement to the properties of native vowels than to those of consonants. However, other types of evidence point to more fundamental differences between consonants and vowels in their linguistic functions and in the control parameters for their production (see, e.g., Fowler, 1980), which may offer alternative (or additional) motivations for the developmental difference between the two classes. For example, vowels but not consonants serve as syllable nuclei; speech errors can occur among vowels or among consonants but not between the two classes; and both speech rate and stress-timing variations affect articulatory movements and muscle control in opposite ways for consonants versus vowels. Generally, vowels are produced as relatively slow alternations in the global shape of the vocal tract, involving primarily the repositioning of the tongue body in the mouth via the larger extrinsic tongue musculature. Consonants are produced as faster, more complex, and more

precisely-timed gestures, involving transient positional changes not only in the tongue body but also the lips and tongue tip (and other articulators), via the additional control of smaller intrinsic articulatory muscles. These articulatory facts result, obviously, in class differences in the temporal and spectral properties of the acoustic signal corresponding to consonants and vowels. Further research will be needed to explore the extent to which developmental differences in the perception of vowels versus consonants may reflect differences in memory for their acoustic properties, as compared to differences in their articulatory properties and/or functional, linguistic roles.

In this context, it would also be of interest to examine how the neural substrate for speech perception may be involved in the development of language-specific influences on perception of vowels and consonants. Both adult and infant dichotic listening studies have found a stronger right ear (RE), or left hemisphere (LH), perceptual advantage for consonants than vowels. The fact that certain acoustic and task manipulations can increase the RE/LH advantage for vowels suggests that the class difference may be attributable to differential loss of auditory information for consonants as compared to vowels during interhemispheric transfer (e.g., Best, 1978; Studdert-Kennedy & Shankweiler, 1970). That is, the underlying LH specialization may be responsive to both consonant and vowel characteristics. But which characteristics are the crucial ones? One study of cross-modal speech perception in infants suggests that the LH is specialized to detect the dynamic gestural properties of speech. MacKain, Studdert-Kennedy, Spieker & Stern (1983) showed that 5-6 month olds recognize which of two adjacent, synched videos of different disyllabic utterances matches the audio of one of the disyllables, but do so only when the matching video is on their right. This attentional bias indicates preferential activation of the LH, according to Kinsbourne's (1978) model of attentional asymmetries. LH specialization for detection of linguistically relevant gestures, rather than simply for processing certain acoustic properties of speech, is supported by evidence that it extends to the rapid manual gestures of sign language [ASL] (e.g., Corina, Vaid & Bellugi, 1992). Research is needed, however, to test for LH sensitivity to the linguistic and gestural differences of vowels vs. consonants (or ASL hand shapes vs. movements), and to examine LH involvement in early phonological development.

Acknowledgments

Work supported by NIH grants DC00403, DC00045, and HD01994. I thank my collaborators on the findings presented here Jerry McRoberts, Nomathemba Sithole Shepherd, Eliza Goodell, Glenda Insabella, and Winifred Strange. Thanks are also extended to colleagues for comments about the perceptual assimilation model (of course, they share no blame for any shortcomings!) Michael Studdert-Kennedy, Carol Fowler, Andrea Levitt, Alice Faber, Janet Werker, Linda Polka, Jim Flege.

References

- Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception. In G. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (eds.) *Child phonology, Volume 2: Perception*. New York: Academic Press
- Best, C. T. (1978). The role of consonant and vowel acoustic features in infant cerebral asymmetries for speech perception. Unpublished doctoral dissertation, Michigan State University.

- Best, C. T. (1984). Discovering messages in the medium: Speech and the prelinguistic infant. In H. E. Fitzgerald, B. Lester, & M. Yogman (eds.) *Advances in pediatric psychology. Volume 2*. New York: Plenum.
- Best, C. T. (1988). The emergence of cerebral asymmetries in early human development: A literature review and a neuroembryological model. In D. L. Molfese & S. J. Segalowitz (eds.) *Brain lateralization in children*. New York: Guilford Press.
- Best, C. T. (1990). Adult perception of nonnative contrasts differing in assimilation to native phonological categories. *Journal of the Acoustical Society of America*, 88, S177.
- Best, C. T. (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. (in press). The emergence of native–language phonological influences in infants: A perceptual assimilation model. To appear in H. C. Nusbaum & J. Goodman (eds.) *The development of speech perception: The transition from speech sounds to spoken words*. Cambridge MA MIT Press.
- Best, C. T., McRoberts, G. W., & Sithole, N. M. (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, 345–360.
- Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross–language perception of approximants. *Journal of Phonetics*, 20, 305–330.
- Brady, S. A., Shankweiler, D., & Mann, V. A. (1983). Speech perception and memory coding in relation to reading ability. *Journal of Experimental Child Psychology*, 35, 345–367.
- Browman, C., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, 3, 219–252.
- Browman, C., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, 6, 201–251.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.
- Cohn, A. (1990). Phonetic and phonological rules of nasalization. *UCLA Working Papers*, 76, May.
- Corina, D. P., Vaid, J., & Bellugi, U. (1992). The linguistic basis of left hemisphere specialization. *Science*, 255, 1258–1260.
- Crowder, R. G. (1971). The sound of vowels and consonants in immediate memory. *Journal of Verbal Learning and Verbal Behavior*, 10, 587–596.
- Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, 51, 419–439.
- Flege, J. E. (1990). Perception and production The relevance of phonetic input to L2 phonological learning. In C. Ferguson & T. Huebner (eds.) *Crosscurrents in second language acquisition and linguistic theories*. Philadelphia: John Benjamins.
- Fourakis, M., & Port, R. (1986). Stop epenthesis in English. *Journal of Phonetics*, 14, 197–221.
- Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, 8, 113–133.
- 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. *Ecological Psychology*, 1, 145–160.
- Fowler, C. A. (1991). Sound–producing sources as objects of perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America*, 88, 1236–1249.
- Fowler, C. A., Best, C. T., & McRoberts, G. W. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception & Psychophysics*, 48, 559–570.
- Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: Cross–modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 816–828.

- Fujisaki, H., & Kawashima, T. (1970). Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Institute* (University of Tokyo), 29, 207-214.
- Insabella, G., & Best, C. T. (1990). Four-year-olds' perception of nonnative contrasts differing in phonological assimilation. *Journal of the Acoustical Society of America*, 88, S177.
- Jusczyk, P. W., & Bertoncini, J. (1988). Viewing the development of speech perception as an innately guided learning process. *Language & Speech*, 31, 217-238.
- Keating, P. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, 60, 286-319.
- Kinsbourne, M. (1978). The biological determinants of functional bisymmetry and asymmetry. In M. Kinsbourne (ed.) *Asymmetrical functions of the brain*. New York: Cambridge University Press.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255, 606-608.
- Lieberman, A. M., Cooper, F. S., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431-461.
- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1-36.
- Lieberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Y. Liberman (eds.) *Phonology and reading disability: Solving the reading puzzle*. Ann Arbor MI: University of Michigan Press.
- Lindau, M. (1982). Phonetic differences in glottalic consonants. *UCLA Working Papers in Phonetics*, 54, 66-77.
- MacKain, K. S. (1982). On assessing the role of experience on infants' speech discrimination. *Journal of Child Language*, 9, 527-542.
- MacKain, K. S., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left hemisphere function. *Science*, 219, 1347-1349.
- Macken, M. (1979). Developmental reorganization of phonology: A hierarchy of basic units of acquisition. *Lingua*, 49, 11-49.
- McCune, L., & Vihman, M. (1987). Vocal motor schemes. *Papers and Reports in Child Language Development*, 26.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.
- Menn, L. (1983). Development of articulatory, phonetic, and phonological capabilities. In B. Butterworth (ed.) *Language production, Volume 2: Development, writing and other language processes*. New York: Academic Press.
- Menn, L. (1986). Phonological units in beginning speech. In A. Bell & J. Hooper (eds.) *Syllables and segments*. Amsterdam: North Holland.
- Menyuk, P., Menn, L., & Silber, R. (1986). Early strategies for the perception and production of words and sounds. In P. Fletcher & M. Garman (eds.) *Language acquisition*. Cambridge UK: Cambridge Univ. Press.
- Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, 46, 505-512.
- Mohanan, K. P. (1986). *The theory of lexical phonology*. Boston: D. Reidel Publishing Company.
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13, 253-260.
- Stevens, K. N. (1968). Acoustic correlates of place of articulation for stop and fricative consonants. *Quarterly Progress Report* (Research Laboratory of Electronics, M. I. T.), 92, 199-205.
- Studdert-Kennedy, M., & Shankweiler, D. (1970). Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, 48, 579-594.
- 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.

- Vihman, M., & Velleman, S. (1989). Phonological reorganization: A case study. *Language & Speech*, 32, 149-170.
- Walton, G., & Bower, T. G. R. (in press). A modal representation of speech in infants. *Infant Behavior and Development*.
- Werker, J. F. (1989). Becoming a native listener. *American Scientist*, 77, 54-59.
- Werker, J. F., & 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., Gilbert, J. H. V., Humphrey, K. & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, 52, 349-355.
- Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, 24, 672-683.