Infant Perception of Non-Native Consonant Contrasts that Adults Assimilate in Different Ways[*]

Catherine C. Best[1] and **Gerald W. McRoberts**[2]

[1] Wesleyan University and Haskins Laboratories

[2] Lehigh University and Haskins Laboratories

Abstract

Numerous findings suggest that non-native speech perception undergoes dramatic changes before the infant's first birthday. Yet the nature and cause of these changes remain uncertain. We evaluated the predictions of several theoretical accounts of developmental change in infants' perception of non-native consonant contrasts. Experiment 1 assessed English-learning infants' discrimination of three isiZulu distinctions that American adults had categorized and discriminated quite differently, consistent with the Perceptual Assimilation Model (PAM: Best, 1995; Best et al., 1988). All involved a distinction employing a single articulatory organ, in this case the larynx. Consistent with all theoretical accounts, 6–8 month olds discriminated all contrasts. However, 10–12 month olds performed more poorly on each, consistent with the Articulatory-Organ-matching hypothesis (AO) derived from PAM and Articulatory Phonology (Studdert-Kennedy & Goldstein, 2003), specifically that older infants should show a decline for non-native distinctions involving a single articulatory organ. However, the results may also be open to other interpretations. The converse AO hypothesis, that non-native *between*-organ distinctions will remain highly discriminate to older infants, was tested in Experiment 2. using a non-native Tigrinya distinction involving lips versus tongue tip. Both ages discriminated this between-organ contrast well, further supporting the AO hypothesis. Implications for theoretical accounts of infant speech perception are discussed.

Keywords

articulatory phonology; cross-language; infant speech perception; non-native consonants; perceptual assimilation

Introduction

Adults have difficulty discriminating many consonant distinctions that are not contrastive in their own languages. Yet young infants show no such language-specific biases during the first half-year (e.g., Aslin & Pisoni, 1980b; Lasky, Syrdal-Lasky, & Klein, 1975; Trehub, 1976; cf. Eilers, Gavin, & Wilson, 1979; Streeter, 1976). Infants under 6–8 months of age can

discriminate both native and non-native consonant contrasts, while infants over 10 months apparently have difficulty discriminating non-native consonants that adult speakers in their language environment have difficulty with (see reviews by Best, 1994b; Werker, 1989). This developmental pattern has been found for English-learning infants tested with the Hindi unaspirated dental versus retroflex stop contrast [t̪] | [ʈ] and voiceless aspirated versus breathy voiced dental stop contrast [t̪ʰ] [d̪ʰ], and with the Nthlakampx (southwest Native Canadian language) velar versus uvular ejective contrast [k'|–[q'], none of which are phonologically distinctive in English (Best, McRoberts, LaFleur, & Silver-Isenstadt, 1995; Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Lalondc, 1988; Werker & Tees, 1984a). Japanese-learning infants show a similar developmental trend for English [ɹ]–[l], which is noncontrastive in Japanese (Tsushima, Takizawa, Sasaki, Shiraki, Nishi, Kohno, Menyuk, & Best, 1994). English-learning infants likewise show a decline in discrimination of a non-native Mandarin fricative-affricate contrast, which older Mandarin-learning infants continue to discriminate (Tsao, Liu, Kuhl, & Tseng, 2000). Similarly, English-learning infants have greater difficulty than Spanish-learning infants with the Spanish alveolar tap versus trill distinction [r]–[r] (Eilers, Gavin, & Oiler, 1982).

The explanation for this early developmental shift remains uncertain, however. Whether, and how, the onset of word comprehension may relate to changes in discrimination of non-native speech contrasts is not yet understood. The average 10-month-old does not yet produce any words, and has just begun to show word comprehension (Benedict, 1979). However it is that early word comprehension may contribute to speech perception, by 10 months infants evidently have discovered some important properties of native consonants, and this has begun to constrain their perception of unfamiliar non-native ones. But just what have they learned, and why does it limit non-native speech perception? Has something changed regarding the type of information they perceive in speech? Is there, for example, a developmental shift from detecting surface acoustic or articulatory details, to perceiving phonetic or phonological information, that is, linguistically relevant information?

Insights may be gained by examining several findings that cHallénge the general claim of decline in discrimination of non-native contrasts by 10–12 months. Older infants apparently do not perceive all non-native contrasts in the same way, that is, discrimination does not always decline developmentally. The earliest report of an alternative developmental pattern found that isiZulu[1] dental versus lateral click consonants [|]–[||] arc discriminated quite well not only by English-speaking adults, but also by infants at least through 14 months, the oldest age tested (Best, McRoberts, & Sithole, 1988). That is, there does not appear to be a developmental decline for this non-native contrast. A follow-up study (Best et al., 1995) reaffirmed that 10–12 month olds discriminate the clicks, even though the same infants failed on the Nthlakampx ejectives for which Werker and colleagues (1984a) had found a performance decline by 10 months. And another lab recently found that both English-and French-learning infants display fairly poor discrimination of the interdental fricative versus alveolar stop distinction [ð]–[d], which is phonologically contrastive in English but not in French (Polka, Colantonio, & Sundara, 2001). Performance by 6–8 and 10–12 month olds from each language environment was comparable to French-speaking adults' relatively poor discrimination, rather than approaching the ceiling-level performance of English-speaking adults as expected. Thus, there appears to be developmental *improvement* in perception of these consonants sometime *after* 12 months of age, but only if the language environment employs them contrastively.

So why do the developmental patterns differ across non-native consonant contrasts? Few accounts of infant speech perception have addressed this variability directly. Most assume a

[1]The name isiZulu refers to the native language spoken by the Zulu people. In prior reports, the name Zulu was incorrectly used to describe speech stimuli recorded in the language isiZulu.

universal pattern of developmental change (see, e.g., Jusczyk, 1986; Jusczyk, 1993; Jusczyk & Bertoncini, 1988; Kuhl, 1993; Werker, 1989; Werker & Pegg, 1992). However, two broader models do predict perceptual differences among non-native contrasts. One posits that non-native contrasts vary along a "fragile-robust" perceptual dimension (Burnham, 1986). Fragile contrasts are defined as distinctions that are low in acoustic salience and rare across the world's languages. Discrimination of these is predicted to decline in the first year if the language environment does not contrast them. Robust contrasts, on the other hand, involve highly salient acoustic distinctions, are common across languages, and show good discrimination until the early school years even without specific experience.

A difficulty with making fragile-robust predictions, however, is that it is not always clear how to determine the acoustic salience level of a given contrast while avoiding circularity (see Best et al., 1988; Polka et al., 2001). Moreover, defining fragility-robustness in terms of both rarity and psychoacoustic salience can be problematic, and certain findings are inconsistent with model's predictions. For example, click consonant contrasts are quite rare across languages. While the psychoacoustic properties of some clicks are presumably quite salient (palatal, alveolar, and lateral clicks), the properties of others are less salient (dental clicks) or even fairly weak (bilabial clicks). More critically, place of articulation contrasts appear to be low in perceptual salience, compared to voicing or manner contrasts (Miller & Nicely, 1955). Thus, click place of articulation contrasts would be fragile on both counts. Yet English-learning infants discriminate the isiZulu dental versus lateral clicks, a place distinction, quite well even beyond the first year, as do English-speaking adults; there appears to be no developmental decline for this contrast (Best et al., 1988). Conversely, the English interdental fricative versus alveolar stop contrast [ð] – [d] is also quite rare. But wheareas [ð] is itself low in acoustic salience (Maddieson, 1984), [d] is robust, and the manner contrast of stop versus fricative is fairly salient perceptually (Miller & Nicely, 1955). Moreover, this contrast also involves a place difference (interdental vs. alveolar), which should somewhat improve perceptual salience. Thus [ð]–[d] seems relatively robust on psychoacoustic grounds, but fragile in terms of its rarity. Very young infants show *poor* discrimination of this contrast, even when it is native. Discrimination *improves* with age if [ð]–[d] is contrastive in the native language, but not until after 12 months (Polka et al., 2001). Thus, the relatively more perceptually robust of the two contrasts just discussed ([ð] – [d]) is more difficult for infants to discriminate. And neither shows a language-experience decline in discrimination in year one, as Burnham's model predicts.

A second model that predicts variations in perception is the Perceptual Assimilation Model (PAM) (Best, 1994a, 1994b, 1995; Best et al., 1988), which is of primary concern to the present research. Sparked by their findings with the isiZulu dental versus lateral clicks, Best and colleagues (1988) developed the PAM to account for a wide range of performance on diverse non-native contrasts. Its central premise is that mature listeners have a strong tendency to perceptually assimilate non-native phones to the native phonemes they perceive as most similar. If there is no clear cut similarity to a single native consonant or vowel, the non-native phone may be perceived as falling in between native phonemic categories (i.e., only weak similarity to two or more), as an uncategorizable speech segment. Rarely, the non-native phone may be so dissimilar from anything in the native system that it is not heard as a phonological element at all, instead being perceived as a nonspcech sound. PAM defines "perceptual similarity" within the frameworks of Articulatory Phonology (e.g., Browman & Goldstein, 1989 e.g., Browman & Goldstein, 1990) and the ecological approach to speech perception (Best, 1984, 1994b; Fowler, 1986). Thus, similarity in the PAM focuses on dynamic articulatory information, that is, on the ways in which articulatory gestures (specific active articulators, constriction locations, and degrees of constriction) shape the speech signal. This view differs from the alternative notion that perceptual similarity, indeed speech perception

generally, is derived from source-neutral acoustic features, for example, properties such as psychoacoustic salience.

PAM proposes that discriminability of a non-native distinction will depend on how the listener assimilates the contrasting phones. Non-native consonants that arc both assimilated as equally-good tokens of a single native consonant should be discriminated poorly (Single-Category [SC] assimilation), whereas those that are assimilated to two different native consonants should show near-ceiling discrimination (Two-Category [TC] assimilation). Contrasting non-native consonants that are assimilated to the same native consonant, but that differ in their perceived degree of similarity to it (Category-Goodness difference in assimilation [CG]), will display good discrimination, intermediate between the SC and TC cases. Thus, discrimination should follow the pattern: TC > CG > SC. This prediction was supported in a recent report on American English-speaking adults' perception of three isiZulu consonant contrasts. The isiZulu voiced versus voiceless lateral fricatives, voiceless aspirated versus ejective velar stops, and plosive versus implosive bilabial stops, respectively, showed TC, CG and SC assimilation and excellent versus good versus poor discrimination (Best, McRoberts, & Goodell, 2001).

Other findings are also consistent with some PAM predictions (e.g., Best & Avery, 1999; Harnsberger, 2000; Polka, 1992; Polka et al., 2001). In addition, English-speaking adults' performance on Werker's original stimulus contrasts appears to be compatible with the PAM prediction that discrimination is better for TC or CG assimilation than for SC assimilation. English-speaking adults show low discrimination levels and little benefit of perceptual training for Hindi unaspirated dental versus retroflex stops $[t^h]$–$[\underline{t}^h]$ and Nthlakampx velar-uvular ejectives $[k']$–$[q']$, but substantially better discrimination and training benefits for the Hindi voiceless aspirated versus breathy voiced dental stops $[t^h]$–$[d^h]$ (Tees & Werker, 1984; Werker et al., 1981; Werker & Tees, 1984b). Both members of the first contrast are likely assimilated as relatively good exemplars of English /d/, and both members of the second as non-prototypical exemplars of English /k/, both satisfying the criteria for SC assimilation. The third contrast, however, likely shows TC assimilation to English /t/–/d/ or CG assimilation as good versus poor /t/.

Tees and Werker themselves, however, offered a different explanation, which provides yet a third theoretical account of variation in perception of non-native phonetic contrasts. Specifically, they posited an allophonic account, in which variations in perception of non-native contrasts are the result of differential native allophonic exposure. They reasoned that English /t/ ($[t^h]$ in initial position) and /d/ provide a comparable voice onset time (VOT) distinction to Hindit $[t^h]$–$[d^h]$, thus offering relevant allophonic experience, but that English provides little to no allophonic exposure to retroflex stops or to velar and uvular ejectives.

The allophonic account is weakened by several findings, however. For one, English listeners discriminate isiZulu clicks quite well, without training, despite the fact that they do not occur allophonically in English (Best et al., 1988). Conversely, English listeners have difficulty discriminating Spanish prevoiced versus short-lag unaspirated stops (e.g., [b] versus [p]), despite the fact that English presents both as allophones of voiced stops (e.g., /b/) (MacKain, 1982). And Werker herself (Pegg & Werker, 1997) reported more recently that while English-learning 6–8 month olds discriminate between English [d] (an initial /d/) and unaspirated [t] (a /t/ following an /s/), both of which occur as allophones of English /d/, 10–12 month olds fail to discriminate.

Still, accounts of variations in adults' perception of diverse non-native contrasts may not apply directly to infants, who have not yet established usage of the native language or its phonology. Children take several years to learn the basic structures of their native language. Full use of the native phonological system—its contrastive functions, phonotactic patterns, and contextual

allophonic variations—is not achieved until the early school years (see, e.g., Ferguson, Menn, & Stoel-Gammon, 1992). Shifts in infants' perception of non-native contrasts presumably reflect the state of their emerging knowledge about native speech. Prior to six months, infants discriminate both native and non-native phonetic contrasts, suggesting that they do not yet recognize native phonemes or contrasts as such.

PAM posits that young infants are simply detecting universal (language-neutral) articulatory patterns in both native and non-native speech. When infants begin to show language-specific effects in discrimination of some non-native contrasts during the second half-year, PAM posits that they have begun to recognize familiar articulatory patterns in native speech, due to perceptual learning or attunement (consistent with the principles of perceptual learning discussed in Gibson & Gibson, 1955). However, this shift does not yet involve recognition of truly phonological information (i.e., segmental elements of an organized system of minimal contrasts). Such phonological ability is likely to be associated with some critical level of lexical and/or morpho-phonemic development, and may not be complete until at least 5–6 years of age (Best, 1993, 1994b). Compatible with the notion that the emergence of contrastive phonology cannot account for the 10–12 month decline in non-native speech discrimination, recognition of familiar words at this age appears to be phonologically underspecified (Hallé & de Boysson-Bardies, 1994; 1996). That is, infants seem to recognize familiar words even when they are "mispronounced" with a phonetic feature change on the initial consonant (i.e., minimal contrast). By 14 months, infants do show finer-grained phonetic representation of familiar words, in that they respond differently to familiar words that are correctly pronounced versus mispronounced (Fennell & Werker, this volume; Swingley & Aslin, 2002; see also Swingley, this volume). But toddlers do not show sensitivity to native minimal contrasts in learning *new* words (artificial-language) prior to about 17–18 months (Stager & Werker, 1997), when the average child begins to produce simple morphology and syntax, display a spurt in vocabulary growth, and show systematic phonologically-motivated patterns in word production.

These observations raise the question of whether infant perceptual development may vary for non-native contrasts that adults assimilate differently. Although there are developmental variations among non-native and even native contrasts, it is not clear whether and how those differences relate to adult patterns of assimilation and discrimination. Nor is it clear what they indicate about the state of the infant's perceptual learning of native speech properties. Whereas the fragile-robust hypothesis and the allophonic-experience hypothesis predict developmental differences for varying types of non-native contrasts, several findings have cast doubt on those predictions, as reviewed above. PAM, on the other hand, predicted the 10–12 month decline in discrimination for non-native contrasts that adults assimilate to a Single Category (SC), as well as the lack of developmental decline for perception of nonassimilable (NA) click consonants. If PAM predictions for adult non-native speech perception extend to other types of non-native contrasts, we should expect a similar lack of decline for Two-Category (TC) assimilations. However, the latter assumption may not be reasonable for infants, and PAM's predictions regarding developmental changes in infants' discrimination of Category Goodness differences (CG), or Uncategorized non-native phones, may also differ from adults' performance patterns. Thus, it would be important to evaluate developmental changes in 6–12 month-old infants' perception for a set of non-native contrasts that adults assimilate and discriminate differently.

If infants do not perceive native consonant distinctions as minimally contrastive elements within a phonological system, as suggested by the research on lexical development and learning summarized above, until about 17–18 months (or possibly as early as 14 months, for familiar words) (see Fennell & Werker, this volume; Hallé & de Boysson-Bardies, 1994; Hallé & de Boysson-Bardies, 1996; Stager & Werker, 1997; Swingley, this volume; Swingley & Aslin,

2002),[2] then their perception of various types of non-native contrasts at 10–12 months must certainly deviate in some ways from adults' assimilation patterns. Consistent with this notion, even though adult English speakers discriminate Hindi [ʈʰ]–[ḍʰ] better than Hindi [ḍʰ]-[ɖ] and Nthlakampx [k']–[q'], English-learning infants showed an equivalent and simultaneous decline in discrimination of all three non-native contrasts by 10–12 months (Tees & Werker, 1984; Werker et al., 1981; Werker & Tees, 1984a, 1984b). In addition, Polka and colleagues (2001) found poor discrimination in English-learning infants throughout the first year for a native contrast that adults discriminated at ceiling. A full understanding of the basis for the perceptual change around 10 months, however, requires systematic developmental comparisons of infants' discrimination for several non-native contrasts on which adults show a wide range of performance. Developmental changes for non-native contrasts on which adults have shown excellent versus good versus poor discrimination would be especially informative. The TC, CG, and SC assimilation patterns in American adults' perception of the three isiZulu consonant contrasts described earlier (Best et al., 2001) satisfy this requirement. Therefore, we tested 6–8 and 10–12 month-old American infants on those three isiZulu contrasts in Experiment 1.

To provide a foundation for predictions about younger versus older infants' perception of these contrasts, we will summarize the adult perceptual findings for each stimulus contrast, the articulatory and acoustic properties of each consonant, and allophonic and nonspeech listening experience that might be relevant. With respect to perceptual findings, the lateral fricative voicing contrast /ɬ/–/ɮ/ (phonetically realized as [ɬ] – [/ɮ]) elicited TC assimilation to an English phonological contrast by adults, who showed near ceiling discrimination (Best et al., 2001). The articulatory difference involved in the contrast is a laryngeal (glottal) gesture distinction involving vocal fold abduction for the voiceless fricative but not the voiced one, a distinction that is also found in English fricative voicing distinctions such as /s/–/z/. Lingual articulation for the lateral fricatives is similar to English /l/ (i.e., constriction of tongue tip and tongue dorsum, such that air flows laterally over the sides of the tongue). However, lateral fricatives involve more constriction along the sides of the tongue than in /l/, resulting in noisy turbulence (i.e., frication). Thus, in the lateral fricative contrast, it is the supralaryngeal articulatory organization that is non-native to English. As for acoustic properties, 24 measurements were made on the multiple tokens of each consonant. Systematic acoustic differences between the lateral fricatives were found on three other measures besides the obvious voicing difference: frication duration was longer and F0 and F1 frequencies at vocalic onset were higher in the voiceless lateral fricatives, consistent with acoustic differences between English voiced and voiceless fricatives (e.g., Pirello, Blumstein, & Kurowski, 1997; Slis & Cohen, 1969). All other acoustic measures showed either partially or completely overlapping ranges of values (for further stimulus details, see Best et al., 2001). We note, however, that voicing distinctions for fricatives as well as stops have been shown to be perceptually robust (Miller & Nicely, 1955). As for English speakers' experience with lateral fricatives, they do not occur as allophones in standard English. While /θ/ and /ð/ preceding /l/ (voiceless /θ/: ATH_L_ETE; voiced /ð/: BLITHE_L_Y) may seem similar to isiZulu lateral fricatives, their tongue tip contact is dental rather than alveolar and they lack lateral frication as in [ɮ] – [/ɬ] In any case, interdental fricative +/l/ sequences are quite rare in English. No nonspeech listening experience appears to be relevant to perception of the lateral fricatives.[3]

The perceptual findings on the isiZulu voiceless aspirated versus ejective velar stop contrast /k/–/k'/ (realized as [kʰ]–[k']) are that it was assimilated by adults as a CG difference in goodness

---

[2]However, for an account of how prelexical infants may use distributional information about allophonic and nonallophonic variants of native phonemes to begin learning something about phonemes, see Peperkamp and Dupoux (2002).

[3]While there is a remote possibility that severe articulatory problems or slurring could yield lateralized productions of the English (intcr) dcntal, alveolar or palatal fricatives /θ/–/ð/, /s/–/z/ or /ʃ/–/ʒ/, exposure to such speech should be quite rare.

of fit to English /k/. It was discriminated quite well, but significantly less well than the lateral fricatives. In terms of articulatory properties, supralaryngeal articulation of both velars is virtually identical to English /k/ when it is aspirated ([k^h]); it is the laryngeal distinction that is non-English. isiZulu and English [k^h] involve the same glottal opening gesture, but the glottal closure for ejective [k′] is not employed contrastively in English. Acoustic measures showed that the release bursts differed systematically between the isiZulu velars: amplitude was higher, duration was longer, and mean weighted frequency was higher at early, mid and late portions of the burst for the ejective than for the voiceless aspirated items. As for allophonic experience with these phones, English provides much exposure to [k^h], but none to ejective stops. With respect to perception of nonspeech properties of the velar stops, Americans typically hear the ejective glottal gesture in isiZulu [k′] as a nonspeech vocal tract event superimposed on a /k/, for example, choking, gagging, throat-clearing, clicking, clacking, clucking, gurgling (Best et al., 2001).

The perceptual findings for the third contrast, the isiZulu plosive versus implosive bilabial stop contrast /b/–/ɓ/(phonetically realized as [p]–[b]), are that the majority of adults showed SC assimilation of both as equally-good /b/'s, and discriminated them rather poorly, though above chance. Articulatorily, the supralaryngeal gesture for the bilabials is identical to English /b/; it is the laryngeal (glottal) distinction that is noncontrastive in English. The glottal setting is virtually identical for isiZulu plosive /b/ and one of the two primary allophones of English /b/: onset of voicing is simultaneous with biliabial release (i.e., voiceless unaspirated [p]) for isiZulu /b/ (Doke, 1926), as well as for voiceless unaspirated allophones ([p]) of English /b/. As for implosive /ɓ/, although older sources state that it is produced with rapid larynx-lowering resulting in negative oral airflow at closure release (Canonici, 1989; Doke, 1926; Maddicson, 1984; Poulos & Bosch, 1997; Van Wyck, 1979; Ziervogel, Louw, & Taljaard, 1985), more recent data indicate that it is no longer realized as an implosive but rather as a prevoiced plosive stop (Giannini, Pettorino, & Toscano, 1988; Traill, Khumalo, & Fridjhon, 1987). The plosive characterization seems appropriate for our isiZulu /ɓ/ stimuli, which are prevoiced and have prominent noise bursts at release (see Best et al., 2001).[4] Thus, isiZulu /ɓ/ is apparently realized phonetically as prevoiced, plosive [b], which is the other of the two primary allophonic variants of English /b/. In terms of acoustic measures, there were four systematic differences between the isiZulu bilabials: release burst amplitude was higher and F0 and Fl onset frequencies were higher, and voice onset time (VOT) was negative (prevoiced) for /ɓ/ ([b]), but short-lag unaspirated for /b/ ([p]). Based on these observations, English allophonic experience would be expected to be ample for both isiZulu /b/ ([p]) and /ɓ/ ([b]). As for nonspeech perception of this contrast, only /ɓ/ evoked nonspeech percepts, which were much less frequent and subtler than for the ejective velars, for example, pursed lips, "harder" pronunciation, or tenser speech muscles (Best et al., 2001).

---

[4]Van Wyck (1979) gives a clear description of /ɓ/ as an implosive, stating (original in Afrikaans; thanks to Prof. Albert Kotzé, University of South Africa al Pretoria, for the translation):

… the nasal cavity and the lips are closed, the vocal cords arc held together lightly and the larynx is lowered to allow pulmonic air to seep through the vocal cords in order to fill the partial vacuum that arises above the glottis. This action causes the vocal cords to vibrate but the vibrations can be sustained for a brief period only because the air pressure above and below the glottis is quickly equalized. … The difference between implosive /ɓ/ and plosive /b/ is that the larynx is lowered much more rapidly for the former than in the case of the latter … [causing] the lower supraglottal air pressure that precedes the release of implosivcs.

However, recent findings indicate that the isiZulu implosive no longer involves rapid larynx-lowering and no longer displays negative airflow al release (Giannini et al., 1988). In addition, the acoustic properties of isiZulu /b/ (Traill, personal communication; Best et al., 2001) appear to be consistent with plosive rather than implosive release (thanks to Nick Clements for pointing this out). Also, isiZulu /ɓ/ causes tone-lowering, a phenomenon that is associated with plosives but not implosives (Traill et al., 1987). The possibility of a historical change from implosive to voiced plosive is consistent with observations that larynx-lowering can vary in a gradient manner, hence resulting in a gradient of ingressiveness to egressiveness (Ladefoged & Maddieson, 1996; Roux, 1991; see also a discussion of the linguistic function of air pressure differences in plosive and nonplosive stops: Clements & Osu, 2002).

Many thanks are due to Nick Clements, Tony Traill, and Albert Kotzé for their helpful input about the phonetic properties of the implosives.

To summarize, in each isiZulu contrast, both members involve identical gestures of the same supralaryngeal articulator(s) and differ by a minimal distinction in articulatory gestures made by a single articulatory organ, the larynx. In other words, none of the contrasts is based on distinction between gestures of different vocal tract articulators, as would be the case with, for example, /p/–/t/ (closure/release gesture of lips vs. tongue tip). Let's review the key properties of the laryngeal contrasts, for purposes of making predictions from various theoretical models. Only one laryngeal distinction is phonologically contrastive in English: that for voiced versus voiceless fricatives. As for the other two non-native laryngeal contrasts, the ejective gesture of the velar contrast is obviously non-English and has a marked effect on the aerodynamics/ acoustics of the velar stop release, whereas both bilabials involve laryngeal settings that occur in English but as noncontrastive allophonic variants of a single English phoneme (/b/. To the extent that allophonic experience may affect perception of non-native contrasts, English allophonic experience is extensive for one member of the isiZulu velar contrast, but is lacking for its cognate. It is weak to nonexistent for both lateral fricatives. And it is substantial for both bilabial stops. Nonspeech qualities appear to contribute substantially to adults' perception of the ejective velar stop, much less to perception of the implosive bilabial /ɓ/ (prevoiced [b]), and not at all to perception of the lateral fricatives.

What predictions may be made about developmental changes in infants' perception of the three isiZulu contrasts, particularly between 6–8 months and 10–12 months, when discrimination has declined for a number of non-native consonant contrasts? A variety of theoretical views offer a range of scenarios. One set of approaches to infant speech perception may be broadly grouped by their common assumption that general information-handling mechanisms, rather than specialized linguistic ones, are responsible for infants' perception of speech. These mechanisms are modified or adjusted by auditory-acoustic experience, as opposed to specifically linguistic forces.

Burnham's (1986) fragile-robust hypothesis falls within this view. Though that proposal has been undercut by several findings, it may still be useful to attempt to locate the three isiZulu contrasts along the fragile-robust dimension, and to make predictions for the present study. Although the lateral fricative voicing contrast is quite infrequent in the world's languages (Maddieson, 1984), voicing contrasts are generally quite perceptually salient (Miller & Nicely, 1955), as summarized earlier. Thus, this isiZulu contrast is probably robust on psychoacoustic grounds alone. By comparison, ejective stops are more frequent across languages, and their most common place of articulation is velar. Additionally, the languages that use ejective velar stops frequently contrast them with the homorganic voiceless plosive, thus /k'/–/k/ ([kʰ]) occurs more frequently than /ɮ/–/ɬ/ (Maddieson, 1984). Given the aerodynamic/acoustic effects of ejective release, the velar stop contrast is also likely to be relatively salient in a psychoacoustic sense. The isiZulu bilabial contrast /b/–/ɓ/ appears to be realized, as summarized above, as voiceless unaspirated [p] versus voiced [b], a highly frequent contrast in the world's languages (substantially more frequent than the English voiced/unaspirated vs. voiceless aspirated contrast) (Maddieson, 1984). Regardless of whether the bilabial contrast is truly plosive versus implosive, or voiceless unaspirated versus prevoiced, both types of distinction are psychoacoustically salient according to Burnham (1986). Thus, all three contrasts appear to be robust. According to the fragile-robust proposal, then, all three isiZulu laryngeal contrasts should still be discriminated well past 10–12 months, declining only later in early childhood.

A number of other general-mechanism accounts have been proposed, including auditory experience-based tuning of sensorineural, psychoacoustic, or attentional mechanisms (e.g., Aslin & Pisoni, 1980b; Kuhl, 1993; Tees & Werker, 1984; Werker & Tees, 1984a, 1984b; Werker et al., 1981; see also Harnsberger, 2000; Polka, 1992; Polka et al., 2001). One such account is Kuhl's Native Language Magnet (NLM) model (1993), which posits that exposure to the acoustic properties of native phonemes results in the formation of phonetic category

prototypes that "warp" the surrounding perceptual space. The prototypes act like perceptual magnets for acoustically similar tokens of the same phonetic category, making the latter difficult to discriminate from them. Nonprototypical members of the category (i.e., poor exemplars) fail to act like magnets, as do nonexperienced non-native phones. Thus, discrimination of acoustically similar tokens is significantly *better* around nonprototypes and non-native phones (i.e., perceptual generalization is poorer) than around prototypes. This suggests that older infants (i.e., 10–12 months) should still be able to discriminate the lateral fricatives well as two nonexperienced nonprototypes, but they should show a modest age-related decline in discriminating the velar stops, which include a native-English prototype ([kʰ]) and a clear nonprototype. They should show a sharp decline in discriminating the bilabials, both of which are prototypical of English /b/.

Cognitive accounts also fall under the general-mechanism rubric. One proposal is that the emergence of certain basic cognitive abilities, which underlie increases in categorization and object search and detour navigation skills at around 10 months, may account for changes in non-native speech category recognition at that age (Diamond, Werker, & Lalonde, 1994; Lalonde & Werker, 1995). Another view is that the developmental change in non-native speech perception is linked specifically to emerging abilities to use correlations among multiple features of experienced stimuli, in order to recognize category identity (e.g., Cohen, 1998; Younger & Cohen, 1983). In the case of speech, infants are assumed to use the multiple features of native phones in order to recognize phonetic category identity. If either cognitive view is correct, discrimination should decline by 10–12 months for the lateral fricatives, which both deviate from familiar phonetic categories. Discrimination should also decline for the bilabial contrast, but in this case because both isiZulu phonemes occur as exemplars of a single native phoneme, such that both should lead to recognition of the same native phonemic category. However, discrimination should remain high for the velar contrast because it involves a comparison between a familiar phonetic category and an unfamiliar one.

In addition to the general-mechanism approaches, however, specialized linguistic accounts also offer developmental predictions for the three isiZulu contrasts. The phonological view posits that truly linguistic, phonemic segments and contrasts emerge in the second half-year. By one such account, young infants display nonlinguistic, psychoacoustic-based perception of speech, but this gives way to perception of linguistic units (i.e., phonemes) when comprehension of word meaning begins to appear around 10 months (WRAPSA model: Jusczyk, 1993; 1994; 1997). Pegg and Werker (1997) offer another phonological account. Finding that English-learning 6–8 month olds, but not 10–12 month olds, discriminate English voiced [d] from unaspirated [t] (both of which occur as allophones of the phoneme /d/), they concluded that the native phonological status of a phonetic distinction governs older infants' perception of noncontrastive native allophones, as well as of non-native contrasts. That is, native contrastive status fosters a phonological reorganization of speech perception at around 10–12 months. The phonological approaches predict that American English-learning infants of 10–12 months should perceive the non-native isiZulu contrasts as do American adults. That is, consistent with PAM findings on non-native speech perception in adults (e.g., Best, 1995; Best et al., 2001), older infants should perceive the lateral fricatives as corresponding to some native phonological distinction and discriminate them better than the velar stops, which they should perceive as a good versus less-good English /k/. They should show great difficulty with the bilabial stops, which they should hear as two good exemplars of English /b/. Note that this view predicts essentially the same pattern of developmental changes across the isiZulu contrasts as does NLM, although the rationale for the predictions is starkly different between the two views.

Phonetic accounts instead posit that although 10–12 month olds have become attuned to phonetic details of the language environment, they do not yet recognize abstract phonological

contrasts per se. One such phonetic view is the allophonic experience hypothesis described earlier (e.g., Tees & Werker, 1984; Werker & Tees, 1984b; see also Maye, Werker, & Gerken, 2002), which would predict a failure in discrimination of the lateral fricatives at 10–12 months due to a dearth of allophonic experience. However, the velar stop contrast should be discriminated, though somewhat less well than at 6–8 months, because it pits an experienced English allophone against a nonexperienced phone. The bilabial contrast should be discriminated quite well, with no decline at 10–12 months, because the phonetic realization of both members of this contrast are frequent allophones of English /b/.

However, the simple allophonic experience hypothesis has been called into question, as noted previously. Werker's current phonetic view instead focuses on the task of word-learning. Recently, she and Stager reported that infants discriminate native minimal contrasts in a pure speech perception task at 14 months, but cannot discriminate the same contrast in a word-learning task until 18–20 months (Stager & Werker, 1997; see also Fennell & Werker, this volume; Swingley, this volume). They concluded that infants under 18 months recognize native phonetic categories but not yet phonological contrasts. It is not entirely clear what this view would predict for our contrasts, though by extrapolation 10–12 month olds should show poor discrimination of the lateral fricatives, neither of which corresponds to a native phonetic category. The bilabials should also be quite difficult for the older infants to discriminate, but because both correspond to the same native phonetic category. There should be only a modest developmental decline in discriminating the velars, which compare a native phonetic category against a non-native one.

PAM presents another phonetic view of early perceptual development, positing that speech perception becomes attuned to native articulatory-phonetic patterns by 10–12 months, specifically to native-language "constellations" of gestures at the segmental or syllabic level (a concept from Articulatory Phonology [AP]: Browman & Goldstein, in preparation; Studdert-Kennedy & Goldstein, 2003). Prior to that attunement, younger infants are assumed to be more universally sensitive to detecting simple differences between single gestures (e.g., tongue tip closure vs. lip closure), rather than noting how gestures are combined into native constellations (e.g., tongue tip closure plus correctly-phased glottal opening for [tʰ] versus [t]). Truly *phonological* attunement, that is, to the native system of minimal contrasts and phonological alternations and morphophonemic patterning, is posited not to be evident until later in development (Best, 1994a, 1994b, 1995). It was the developmental PAM viewpoint that was of primary interest to us here. Because of the articulatory assumptions of PAM, we believed it important to extend the model to make specific predictions based on AP principles regarding articulatory gestures. We did so by including Goldstein's articulatory organ (AO) hypothesis (Browman & Goldstein, in preparation; Studdert-Kennedy & Goldstein, 2003), which he generated to extend the theoretical framework of AP to early development, specifically in order to address how infants learn speech by imitating articulatory gestures (see also Meltzoff & Moore, 1997; Studdert-Kennedy, 2002). The organ hypothesis posits that what infants detect in a speech segment (or syllable/word) is the primary articulatory organ(s) (e.g., lips, larynx, velum) that produced it; infants are posited to be much less likely to recognize the parametric details of the gesture (speed, precise location). Thus, they will have greater difficulty discriminating a minimal phonetic contrast distinguished by two different gestures made by the same primary articulator (i.e., within-organ contrasts), than discriminating a minimal contrast distinguished by a given gesture produced by different articulators (i.e., between-organ contrasts).

We combined the articulatory organ hypothesis with PAM's assumption that infants become attuned to native gestural constellations by the end of the first year. This led to the prediction that discrimination of non-native within-organ contrasts will decline earlier and more dramatically than discrimination of non-native between-organ contrasts. The three isiZulu

contrasts are each within-organ laryngeal distinctions, either involving a non-native laryngeal gesture (velar ejective), a native laryngeal distinction in the context of a non-native supralaryngeal gesture pattern (lateral fricatives), or a laryngeal distinction that occurs but is noncontrastive in the native language (bilabial stops). Thus, according to the PAM/articulatory organ (PAM/AO) hypothesis, 10–12 month olds should show a decline in discrimination of minimal within-organ contrasts between non-native phones that they hear as members of a native phonetic category, which should be the case for both the isiZulu bilabial stops and velar stops. In the case of the lateral fricatives, neither member of this within-organ contrast corresponds to any native phonetic categories, and they differ from one another only by different laryngeal gestures. Thus, 10–12 month olds should also show a decline in discriminating this non-native distinction, even though adults assimilated it as a TC phonological contrast and discriminated it quite well. Thus, by our PAM/OA reasoning, 10–12 month olds should show a decline in discrimination, relative to 6–8 month olds, for all three isiZulu contrasts.

The goal of this report, then, was to evaluate the PAM/AO hypothesis against the other theoretical possibilities described earlier (see Table 1 for summary of predictions). The findings should help to better determine the nature of the changes that occur in native phonological development during the first year. Experiment 1 focused on 6–8 versus 10–12 month olds' discrimination of non-native single-organ (laryngeal) contrasts that American English-speaking adults had assimilated respectively as TC, CG, and SC contrasts. A non-native between-organ contrast, which American adults had also assimilated as a TC contrast, was tested for comparison in Experiment 2.

Experiment 1

2.1 Method

Participants—The final data set included 11 infants at 6–8 months (M_{age} = 7 months 10 days; range = 6 months 8 days to 8 months 10 days) and 11 infants at 10–12 months (M_{age} = 11 months 15 days, range = 10 months 21 days to 12 months 27 days). All were normal, full-term infants without gestational or labor/delivery complications, and were free of ear infections or colds on the day of testing. These infants had all successfully completed three tests within the study session, one for each of the isiZulu stimulus contrasts. An additional 24 infants at 6–8 months and 13 infants at 10–12 months were tested but later excluded from the study for crying, equipment failure, experimenter error, parental interference, or inattentiveness (i.e., 10 or more consecutive trials without visual fixation responses),[5] or because of ear infection/cold on the test day, pregnancy/delivery complications, or familial speech/language disorders.

Stimulus materials—The stimuli were from Best et al. (2001): [ɬ] – [ɮ](voiced vs. voiceless lateral fricatives), [kʰa] – [kⁱa] (voiceless aspirated vs. ejective velar stops), and [pu] – [ɓu] (unaspirated plosive vs. implosive bilabial stops). Different vowels were used for each contrast in order to maintain infants' attention across the three required tests, as this was a within-subjects design.

The syllables all had high tone on the vowel[6] and were spoken by an adult female native speaker of isiZulu from Durban, South Africa. Six tokens of each syllable had been selected from the recordings; the contrasting sets of tokens had been chosen to match as closely as possible on

---

[5]The relatively high rejection rates are partly due to the fact that infants had to successfully complete three habituation tests within a session, on each of the three isiZulu contrasts, in order to be included in the study. The power of within-subject comparisons, especially in light of the predictions being tested, was deemed to be of sufficient importance to outweigh an increase in rejection rate. Other contributing factors were that infants could not be included if they had ear infections, familial language problems, and other conditions that may interfere with speech perception. Often these problems were not identified until after the infant had already been tested.
[6]isiZulu is a tone language that differentiates between high and low tones on syllable nuclei.

all acoustic dimensions other than those critical to the phonetic distinction. (For full details on stimulus development and acoustic measurements, see Best et al., 2001.)

Procedure—We employed the same infant-controlled visual fixation habituation procedure used in our previous studies (Best et al., 1988; Best et al., 1995; see also Miller, 1983). Random tokens of one stimulus category were played to the infant over a hidden loudspeaker at a conversational listening level (65–70 db SPL) whenever the infant fixated on a colored checkerboard directly facing him/her, which was rear-projected onto a sound-attenuating window that separated the test room and the adjacent observation room. Thus, the infant was conditioned to fixate the checkerboard, which was reinforced with speech presentations, analogous to the experimental contingencies in the well-known high-amplitude sucking habituation procedure (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971). A video camera hidden below the checkerboard display was connected to a video monitor in the observation room, allowing the experimenter to monitor the infant's fixations of the checkerboard via corneal reflections and other visible indices of directed gaze (e.g., head/eye orientation). Fixations and bouts of crying or sleeping were recorded via key press from an observer response keyboard connected to a computer that controlled the presentation of audio stimuli from an Otari 5050MXB reel-to-reel tape deck, dependent on the infant's fixation pattern.

Trial duration was under infant control. For as long as the infant fixated the projected checkerboard, audio tokens from one stimulus category were presented (ISI = 750ms). Audio presentations ceased (after completing the ongoing stimulus token) whenever the infant looked away. A given trial continued for as long as the infant maintained fixation, or if the infant returned to fixating after a brief look-away of less than 2 s. However, if the infant looked away for 2 s consecutively, the trial ended and the checkerboard disappeared during the 1 s intertrial interval, after which the checkerboard automatically reappeared, signaling the beginning of a new trial. Habituation was defined as two consecutive trials with fixation durations below 50% of the mean of the two highest preceding trials (Miller, 1983). The habituation criterion was calculated and updated on a trial-to-trial basis by the experimental computer program. Once habituation was met during the first phase (familiarization), audio presentations shifted to the contrasting stimulus category for the test phase, which continued until the infant again met the habituation criterion.

During testing the infant sat in an infant seat or on the parent's lap in a dimly lit 2 m × 1 m × 1 m testing booth, at a distance of approximately .5 m from the rear-projection window. The booth was open at the back and its sides were covered with black fabric. The wall at the front of the booth was also covered with black fabric, except for the 0.6 m × 0.6 m area directly in front of the infant where the checkerboard was projected; a small opening for the video camera lens was below the display. A loudspeaker (Jamo mini-speaker), attached to the wall 1 m above the projection window and hidden behind the black cloth covering, was used for stimulus presentations. Both the parent and the experimenter observing the infants' fixations listened to music over circum-aural headphones (Scnnheiser HD440) during tests to prevent them from hearing the stimuli and inadvertently influencing the infant or the fixation observations.

Each infant completed a discrimination test on each of the three stimulus contrasts within a single session (see also Best et al., 1988, 1995). Test order was randomized across infants within each age group. Short breaks of 5–10 mins were taken between tests if necessary to maintain infants' attention and/or to soothe them if they had become irritable. Otherwise, the session proceeded from one test to the next with just a 1–2mins break to reposition the audio tape and restart the computer program. Infants were eliminated from the final data set if they cried for more than a cumulative 30 seconds during any test, or if they cried during any trials just before or after the test shift.

2.2 Results

Interobserver reliability—The data for a random selection of 16 infants (i.e., 48 individual tests) were rescored by second observers, who reran the testing program while viewing the infants' test session videotapes (i.e., off-line). Interobserver reliability was evaluated statistically via rank-order correlations of the per-trial looking times registered by the original and second observers. Reliability was quite good (M_r = 0.97; range = 0.77 to 1.00).[7]

Discrimination results—Discrimination was assessed by comparing mean fixation duration during the last two trials of the familiarization phase (preshift block) to mean fixation during the first two trials of the test phase (postshift block). The postshift block was defined as beginning with the first trial after the stimulus shift in which the infant fixated on the slide and thus had an opportunity to begin hearing the test stimuli (Best et al., 1988; Best et al., 1995). A significant increase in fixation during the postshift block relative to the preshift block is taken as evidence that infants detected the stimulus change. The data were entered into an Age (6–8 months, 10–12 months) × Stimulus Contrast (lateral fricatives, velars, bilabials) × Trial Block (preshift, postshift) analysis of variance (ANOVA). Test order was not included as a factor because preliminary analyses indicated it had no systematic effect on discrimination. Trial Block was the only significant overall effect: infants looked longer in postshift trials than preshift trials (M = 6.22 vs. 3.49s, respectively), indicating overall recovery of fixation during test trials (i.e., reliable discrimination), $F(1, 40) = 12.01$, $p < .003$.

Age groups—To more directly evaluate a priori hypotheses about non-native speech discrimination performance at each age (Table 1), separate Trial Block × Stimulus Contrast ANOVAs were conducted for each age group. For the 6–8 month olds, the Trial Block effect was significant, $F(1, 10) = 8.413$, $p < .016$, indicating reliable discrimination across all contrasts. However, for the 10–12 month olds the Trial Block effect was nonsignificant, $F(1, 10) = 3.73$, $p > .082$. No other effects approached significance at either age.

Stimulus contrasts—To further examine the a priori predictions (Table 1) about differences between 6–8 and 10–12 month olds' responses to the individual stimulus contrasts, separate Age × Trial Block ANOVAs were conducted for each contrast, followed by simple effects tests on the interaction. This allowed a direct test of age differences in discrimination of each isiZulu contrast.

For the bilabial stop contrast, on which adults had shown SC assimilation and poor discrimination, the infants' Trial Block effect was significant, $F(1, 10) = 9.003$, $p < .008$. Simple effects tests on the Age × Trial Block interaction revealed that there was significant postshift recovery of fixation (i.e., reliable discrimination) by the 6–8 month olds, $F(1, 20) = 6.099$, $p < .025$. But there was nonsignificant recovery by the 10–12 month olds, $F(1, 20) = 3.146$, $p > .09$, indicating that discrimination was unreliable in the older group.

Adults had shown CG assimilation of the velar stop contrast, with good discrimination. Across infant ages, the Trial Block effect for this contrast was significant, $F(1, 20) = 6.637$, $p < .02$. Simple effects tests on the Age × Trial Block interaction indicated that postshift recovery of fixation (i.e.. discrimination) was nearly significant for the 6–8 month olds, $F(1, 20) = 4.051$, $p = .058$, but not in the 10–12 month olds, 20) = 2.659, $p = .12$. No other simple effects approached significance for this contrast.

For the lateral fricatives, which had yielded TC assimilation with excellent discrimination in adults, the infants' Trial Block effect was not significant. Simple effects tests on Age × Trial

---

Block interaction indicated that the younger age showed significant recovery of postshift fixation. $F(1, 20) = 6.46$, $p<.02$, but the older group's performance did not even approach significance ($p > .85$). Moreover, while there was no reliable age difference in preshift fixation levels, the younger group displayed significantly greater postshift fixation than did the older group, $F(1, 39) = 4.368$, $p< .045$. Thus, the younger group discriminated the lateral fricative contrast, whereas the older group did not.

Figure 1 displays performance on each contrast by the two age groups. In sum, there was significant evidence of discrimination by the younger group for the bilabial and lateral fricative contrasts, with nearly significant discrimination for the velar contrast. By comparison, the older group failed to show reliable discrimination of any of the contrasts.

2.3 Discussion

The results of Experiment 1 are consistent with the general expectation of most models reviewed, that 6–8 month olds would discriminate all three isiZulu contrasts, and that 10–12 month olds would show a decline in discrimination for one or more contrasts (see Table 1). In actuality, the 10–12 month olds failed to discriminate not only the bilabial stops, on which adults had shown poor discrimination, but also the velar stops and the lateral fricatives, on which adults had respectively shown good and excellent discrimination (Best et al., 2001). Considering the various hypotheses described earlier, these results are most consistent with PAM/AO, which hypothesized that older infants have become attuned to native phonetic constellations but not to phonological principles, and predicted that 10–12 month olds would show a decline relative to the 6–8 month olds in discriminating non-native within-organ distinctions. PAM/OA is the only hypothesis that predicted 10–12 month olds' difficulties with all three isiZulu contrasts (see Table 1). We note, in addition, that simple psychoacoustic principles do not seem to easily explain the decline in the older infants' discrimination of all three minimal laryngeal contrasts, given that voicing contrasts have been found to be perceptually robust (Miller & Nicely, 1955), and that the ejective versus voiceless aspirated stop contrast and the plosive versus implosive stop contrast (phonetically, a prevoiced vs. voiceless unaspirated) were posited to be psychoacoustically robust (see Burnham, 1986). Certainly, at the very least, the perceptual salience of the fricative voicing contrast should otherwise have made it quite discriminable to the older infants on a psychoacoustic basis.

While the results appear most compatible with the PAM/AO view, however, converging support is needed, especially given that no differences among the three contrasts were predicted for the older infants (or the younger infants, for that matter). Moreover, additional possible explanations of the older infants' difficulties must be considered. Toward those ends, Experiment 2 involved PAM/AO predictions of perceptual differences among non-native contrasts for 10–12 month olds. It focused specifically on teasing apart the factors that may contribute to developmental changes in discriminating the type of contrast represented by the lateral fricatives: those that adults assimilate as TC contrasts (i.e., as phonological distinctions) and discriminate quite well. In Experiment 1, it was the lateral fricative TC contrast that had yielded the most striking difference between older infants' poor performance and prior findings of excellent performance in adults.

Experiment 2

To evaluate whether the older infants' difficulty with the lateral fricatives might be attributable to the fact that these represent within-organ gestural distinctions, Experiment 2 incorporated a between-organ distinction that had also shown TC assimilation and excellent discrimination in American English-speaking adults: the bilabial versus alveolar ejective stop contrast/p′/–/t′/ of Tigrinya, a language spoken in Eritrea (Ethiopia) (Best et al., 2001). Both [p′] and [t′] involve a non-native ejective laryngeal gesture like that in isiZulu [k′]. However, they differ

in their supralaryngeal gestures, which involve two different articulatory organs: lips versus tongue tip. Because [p′]–[t′] is a between-organ contrast, the PAM/OA approach predicts good discrimination by older infants, in contrast to their decline in discrimination of the lateral fricatives relative to younger infants. For comparison, we also included a replication with the isiZulu lateral fricatives. This was a particularly important comparison, given that (1) we wished to utilize a pair of contrasts for which PAM/AO predicts different developmental patterns, and (2) we decided it was important to modify the habituation criterion (see below).

Several other factors besides the articulatory organ difference, in addition, could potentially have contributed to the older infants' difficulty with the lateral fricatives, and needed to be explored. For example, their emerging sensitivity to English phonotactic constraints (see Jusczyk, 1994) may have interfered with their discrimination of the lateral fricative syllables, which violated the English phonotactic rule against open syllables ending in lax vowels such as [ɛ]. To evaluate this phonotactic hypothesis, we used [ɛ] in an open-syllable context for the new between-organ non-native contrast, also. If the violation of native phonotactic constraints was responsible for the 10–12 month olds' difficulty, then they should show an equivalent level of difficulty with the Tigrinya ejectives in this context, contrary to the PAM/OA prediction of good discrimination.

Alternatively, a few researchers have claimed that both older *and* younger infants have difficulty discriminating fricative voicing distinctions, even including native fricative voicing contrasts (e.g., Eilers, 1977; see discussion by Burnham, 1986). The fact that the younger infants in Experiment 1 discriminated the lateral fricative voicing contrast casts some doubt on this hypothesis. Nonetheless, we evaluated it for thoroughness, and to compare it against the PAM/AO prediction that older infants should show a decline in discrimination of within-organ (laryngeal) minimal contrasts, relative to younger infants. Therefore, we also included an English fricative voicing contrast, using the same vowel and syllable context as the lateral fricatives. This contrast employs the same within-organ contrast as the lateral fricatives, yet it occurs in English. Most of the viewpoints listed in Table 1 would predict that because this contrast occurs in the infants' language environment, older infants should continue to discriminate it very well. On the other hand, if there is anything to the claim that infants have difficulty with fricative voicing distinctions, regardless of language experience, then infants of both ages should show a decline in discrimination of this native contrast as well as the non-native lateral fricative voicing contrast. With respect to the phonotactic hypothesis summarized in the preceding paragraph, if older infants' growing sensitivity to native phonotactic violations affects discrimination even for native consonants, they may show a decline in discrimination of all three Experiment 2 contrasts.

See Table 2 for a summary of the properties and various theoretical predictions for the non-native and native contrasts tested in Experiment 2.

3.1 Method

Subjects—The final data set included 15 infants at 6–8 months (M_{age} = 7 months 5 days, range = 5 months 28 days to 7 months 25 days) and 14 infants at 10–12 months (M_{age} = 11 months 4 days, range = 10 months 2 days to 12 months 28 days). Criteria for inclusion were as described in Experiment 1. An additional 48 infants were excluded from the final set (22 at 6–8 months and 18 at 10–12 months).

Stimulus materials—The non-native contrasts were the isiZulu lateral fricative contrast from Experiment 1, and the Tigrinya bilabial versus alveolar ejective stops [p′]–[t′] produced by a male native speaker from Eritrea (Ethiopia). The native contrast was English /s/-/z/ produced by a female native speaker (author CTB) (Best et al., 2001). The native contrast was chosen to be similar, acoustically and articulatorily, to the isiZulu lateral fricatives: a fricative

voicing contrast involving a tongue constriction gesture. Both the Tigrinya and the English syllables employed the same vowel ([E]) as the isiZulu syllables. Thus, any differences in discrimination among the three contrasts would be due to the consonants and not to vowel or phonotactic effects. As before, there were six tokens per category, matched as closely as possible between the contrasting syllables of each pair for overall duration, fundamental frequency and contour, and vowel formant frequencies (see Best et al., 2001).

The primary acoustic differences between the Tigrinya ejectives were in the spectrum, duration, and amplitude of the release bursts. The English fricatives differed primarily in voicing and amplitude of the fricatives and F0 and F1 onset frequencies, comparable to the difference between the isiZulu lateral fricatives.

Procedure—We employed the same procedure as in Experiment 1, except that the habituation criterion was made more stringent in order to assure full habituation, which in turn optimizes the likelihood of response recovery during the test phase. Optimizing the chance of response recovery during the test phase was crucial, given predictions that the older infants should show a decline, relative to younger infants, in discriminating one or more of the contrasts. The three highest-looking trials of the familiarization phase were used to calculate the habituation criterion, rather than only two trials as in Experiment 1. Also, three consecutive trials with looking durations below the habituation criterion (rather than only 2 such trials) were required for the shift to the test phase. We note that this more stringent habituation criterion may actually increase the possibility of spontaneous response recovery, that is, of spurious evidence for discrimination. This observation is particularly relevant to older infants' performance on the two fricative voicing contrasts.[8]

3.2 Results

Interobserver reliability—The data for a random selection of 13 infant subjects (i.e., 39 individual tests) were second-observed from the session videotapes, as in Experiment 1. Rank-order correlations of the per-trial looking times registered by the first and second observers indicated that interobserver reliabilities were excellent (M_r = 0.98, range = 0.91 to 0.99).

Discrimination results—Discrimination was assessed as in Experiment 1, using an Age (6–8 months, 10–12 months) × Stimulus Contrast (Zulu, Tigrinya, English) × Trial Block (preshift, postshift) ANOVA. Test order was not included as a factor because preliminary analyses showed that it did not have any systematic effect on discrimination. Trial Block (Preshift vs. Postshift) was the only significant effect, $F(1, 28) = 50.24$, $p < .0001$. Infants looked longer during postshift trials than during preshift trials (M = 4.08 vs. 0.88 s, respectively). The Age difference in overall looking, averaged across the pre- and post-shift trial blocks, was marginal, $F(1, 28) = 3.370$, $p < .08$. The younger group showed a somewhat higher overall looking level than did the older group. The differences in mean fixation among the three stimulus contrasts, averaged across trial blocks, was also marginal, $F(1, 28) = 2.898$, $p = .06$. Overall, mean looking levels were somewhat higher for the native contrast than for the two non-native contrasts. Figure 2 presents the preshift versus postshift performance of the two ages on the three contrasts.

Age groups—To evaluate hypotheses about overall age differences in non-native speech discrimination, separate Trial Block × Stimulus Contrast ANOVAs were conducted for each age group. The only significant effect was Trial Block, both in the 6–8 month olds, $F(1, 14) = 28.344$, $p < .001$, and in the 10–12 month olds, $F(1, 14) = 23.484$, $p < .001$. Thus, both age groups showed reliable recovery of fixation during the test phase, across stimulus contrasts. This is

---

[8]We thank an anonymous reviewer for pointing this out to us.

presumably associated with the more stringent habituation criterion, that is, to the fact that preshift looking levels were required to be quite low.

Stimulus contrasts—A priori predictions were that 6–8 month olds would discriminate all three contrasts, but 10–12 month olds would show lower discrimination performance on the isiZulu lateral fricatives, possibly the English fricative voicing contrast, and perhaps the Tigrinya contrast. To assess these possibilities directly, separate Age × Trial Block ANOVAs were conducted for each stimulus contrast individually, followed by simple effects tests.

For the Tigrinya place of articulation contrast between ejective stops, on which adults had shown TC assimilation and excellent discrimination, the Trial Block effect was significant, F $(1, 28) = 14.79$, $p < .001$. Simple effects tests on the Age × Trial Block interaction revealed significant postshift recovery of fixation by both the 6–8 month olds, $F(1, 28) = 8.139$, $p < .01$, and the 10–12 month olds, $F(1, 28) = 6.686$, $p < .02$. There were no significant age differences.

For the native English fricative voicing contrast, the Trial Block effect was significant, $F(1, 28) = 27.268$, $p < .0001$. As with the Tigrinya contrast, neither the Age effect nor the interaction was significant. Simple effects tests on the interaction indicated postshift fixation recovery for both the 6–8 month olds, $F(1, 28) = 23.124$, $p = .0001$, and the 10–12 month olds, $F(1, 28) = 6.636)$, $p = .02$. Importantly, however, the simple effects tests indicated that the younger group displayed significantly greater response recovery on the test trial than the older group, $F(1, 56) = 4.038$, $p < .05$. That is, the younger group showed significantly greater dishabituation (i.e., discrimination).

The isiZulu lateral fricatives also yielded a significant Trial Block effect, $F(1, 28) = 35.912$, $p < .0001$. However, in this case the Age effect was also significant, $F(1, 28) = 4.518$, $p < .05$, as was the Age × Trial Block interaction, $F(1, 28) = 6.988$, $p < .02$. Simple effects tests of the interaction indicated that recovery of postshift fixation was achieved by both the younger group, $F(1, 28) = 37.291$, $p < .0001$, and the older group, $F(1, 28) = 5.608$, $p = .025$. Importantly, as with the English contrast, there was no age difference in fixation during the preshift trial block, but the younger group showed significantly higher levels of fixation for the postshift trial block than did the older group, $F(1, 56) = 10.894$, $p < .005$. That is, the older group showed a decline in discrimination relative to the younger group on this and the English contrast, although no such age difference appeared for the Tingrinya contrast.

3.3 Discussion

The results of Experiment 2 appear to rule out the hypothesis that a violation of native phonotactic rules contributes to 10–12 month olds' difficulties with discriminating a non-native consonant contrast. Despite the fact that all three contrasts were presented in the same context of an open syllable ending in the lax vowel [ɛ], which is phonotactically impermissible in English, they discriminated the Tigrinya contrast on a par with 6–8 month olds, and discriminated the Zulu and English contrasts but at significantly lower levels than the younger group. Had this phonotactic violation impaired the older infants' performance, they should have displayed comparable difficulties in discriminating all three contrasts.

Why is it that the non-native Tigrinya contrast was discriminated equally well by both the 6–8 and 10–12 month olds, whereas the older infants discriminated the other two contrasts significantly less well than did the younger infants? Some would wish to consider acoustic properties of the three distinctions as a possible contributing factor to the discrimination pattern (although we must note that a simple acoustic account is unlikely to explain, on its own, why younger infants are unaffected by acoustic differences among the contrasts, whereas older infants are). As for acoustic bases of the three non-native contrasts in Experiment 2, measurements by Best and colleagues (2001) indicate that the ejective bursts are brief ($M = 12$

vs. 20 ms) and differ in spectral energy distribution but not in VOT, whereas both of the fricative distinctions are marked by long-duration consonant noise portions (frication) that differ in voicing and duration, being longer for the voiceless than the voiced cognates (isiZulu: M = 152 vs. 108ms.; English: M = 238 vs. 212ms). The ejectives involve a place of articulation difference, and thus have different formant transition patterns; because the fricative distinctions are in voicing and not place, they are not marked by differences in vocalic transitions. However, because the ejectives are voiceless, their vocalic formant transitions arc truncated, and because they arc ejectives, there is virtually no spectral movement during the glottal closure period between oral release and delayed onset of voicing. In other words, the formant transition differences between the two ejectives are modest, at best. Moreover, recall that consonant voicing differences are substantially more salient perceptually than are place of articulation differences (Miller & Nicely, 1955). In summary, it appears that acoustic properties of the ejective versus the fricative contrasts do not account for the older infants' differences in discrimination of the three contrasts, relative to the younger group.

Another property that distinguishes the Tigrinya contrast from the other two — the property that we believe is a more likely source of the perceptual patterns — is that the bilabial versus alveolar ejective contrast involves a between-organ distinction, whereas the isiZulu and English contrasts are within-organ distinctions. The finding that both ages discriminate the Tigrinya contrast equally well, but that younger infants show significantly stronger discrimination than the older infants for the fricative voicing contrasts is consistent with the PAM/AO hypothesis that discrimination of within-organ minimal distinctions should show more age-related decline than does discrimination of between-organ distinctions. In fact, it appears that this decline may occur not only with non-native, but also with (certain) *native* within-organ distinctions.

That interpretation may need to be qualified, however. Both within-organ contrasts were fricative voicing distinctions, one non-native and the other native. It may be that this specific type of phonetic distinction, rather than within-organ contrasts in general, is responsible for the decline in older infants' discrimination. Additional investigations with other types of within- and between-organ contrasts, both native and non-native, will be needed to determine whether 10–12 month olds' difficulties are specific to fricative voicing contrasts or instead generalize to other within-organ distinctions. For example, it would be important to investigate developmental changes in perception of stop voicing contrasts as within-organ distinctions, in addition to the current fricative voicing contrasts. Although previous studies have investigated perception of non-native stop voicing distinctions (e.g., Aslin & Pisoni, 1980a; Eilers et al., 1979; Eimas et al., 1971; Lasky et al., 1975; Streeter, 1976), none of these tested developmental changes in the second half-year of life, and importantly, none used naturally produced non-native stops. With the synthetic stops used in the prior experiments, only VOT was manipulated, whereas in natural speech phonetic details other than VOT may be crucial to certain stop voicing distinctions in some languages (e.g., Spanish, Korean).

One detail of the findings that needs to be addressed is the fact that the older infants failed to discriminate the lateral fricatives in Experiment 1, but discriminated them in Experiment 2 (while keeping in mind that simple effects tests indicate that even in the second experiment the older infants showed significantly less recovery of fixation during the test trials than the younger infants did). The cross-experiment difference in the older infants' performance on the lateral fricatives was most likely due to the difference in stringency of the habituation criteria in the two experiments. Because the criterion was more stringent in Experiment 2, it may be that the infants were more fully habituated prior to the test trials, a suggestion that is borne out by comparing the preshift fixation levels for the two experiments. This would increase the probability of response recovery during the test phase, even for difficult contrasts. At the same time, it must be noted that a very stringent habituation criterion, and thus very low levels of

looking during the immediate preshift trials in Experiment 2, actually *raises* the possibility of spontaneous response recovery (i.e., false positive evidence of discrimination). If indeed the response recovery during the test trials was spurious, however, we would have greatest cause to be suspicious of whether discrimination was reliable in those conditions where the recovery was more modest relative to the younger infants, that is, for the older infants' discrimination of the non-native and native fricatives — exactly the two cases in which the PAM/AO predicts a decline in discrimination.

General Discussion

Across the two experiments, the 6–8 month olds discriminated all four non-native consonant contrasts, as well as the native fricative voicing contrast, with no performance differences among the five. This was as expected by all of the theoretical accounts discussed earlier (see Table 1). By comparison, according to statistical tests the performance of the 10–12 month olds was significantly weaker than the younger group for all but one non-native contrast, as well as for the native fricatives. The only case in which the older and younger infants did not differ was a non-native contrast, the Tigrinya ejectives [p′]–[t′]. It is the performance of the older group relative to the younger group across these five contrasts that is most informative in evaluating the various theoretical accounts raised in the general introduction (Table 1) and in the rationale for Experiment 2 (Table 2).

The pattern of older versus younger infants' performance over the two experiments deviates in one or more ways from the predictions of all but one of the theoretical models reviewed. With respect to Experiment 1, the fragile-robust hypothesis appears to predict that 10–12 month olds, as well as 6–8 month olds, would show very good discrimination of each of the isiZulu laryngeal contrasts, yet the older infants showed significantly poorer performance on all three contrasts. The older infants' poorer performance on the isiZulu bilabial and velar stop contrasts, relative to the younger infants, is consistent with the auditory tuning hypothesis and the phonological hypothesis. However, those models both predict that 10–12 month olds' performance should be very good for the lateral fricative contrast, which was clearly not the case.The general cognitive hypothesis, conversely, was consistent with the older infants' poor performance on the lateral fricatives and bilabial stops, but was undercut by their decline on the velar stop contrast. The allophonic experience hypothesis predicted the older infants' decline in discriminating the lateral fricatives and velar stops, but its prediction that they would discriminate the bilabial stops was not upheld. The native word phonetics hypothesis is compatible with the older infants' performance decline on the bilabial and velar stop contrasts. And it appears to be in line with the older infants' difficulty discriminating the lateral fricative voicing contrast. However, the older infants' discrimination of the lateral fricatives in Experiment 2, though at a significantly lower level than the younger infants, is at odds with this model's expectations of failure on that contrast by that age group. Moreover, the allophonic and the word phonetics hypotheses would predict a decline or failure of 10–12 month olds on the Tigrinya ejective place contrast in Experiment 2, on which the older infants instead performed on a par with the younger ones.

Also with respect to the Experiment 2 findings, the phonotactic hypothesis (Jusczyk, Luce, & Charles-Luce, 1994) was undermined by the findings that older infants' performance differed across the stimulus contrasts even though all three utilized the same lax vowel [ɛ] in open (CV) syllables, which is impermissible in English. A pure native phonotactic effect should have been constant across all contrasts, and surely would not have been stronger for the native contrast (English fricatives) and one non-native contrast (isiZulu lateral fricatives) than for another non-native contrast (Tigrinya ejectives), as we found in Experiment 2. The third alternative posed in Experiment 2, the fricative voicing hypothesis, is weakened by the fact that the *younger* infants had no difficulty with native or non-native fricative voicing distinctions. Moreover, an

account based on differences in the acoustic properties and relative perceptual salience of the various contrasts tested docs not account for the pattern of performance, as summarized in the Discussion sections for each experiment.

The findings across the two experiments are, however, consistent with the predictions of PAM/ AO: the Perceptual Assimilation Model (Best, 1993; 1994b, 1995; Best et al., 1988) as extended by the articulatory organ hypothesis derived from Articulatory Phonology (Browman & Goldstein, in preparation; Studdert-Kennedy & Goldstein, in press). PAM/AO predicted that 10–12 month olds should show a decline in discrimination of all three isiZulu laryngeal contrasts, despite the fact that discrimination of these contrasts differed greatly for adults from their language environment, from relatively poor for the bilabial stops, to good for the velar stops, to excellent for the lateral fricatives. The PAM/AO account of these differences between the older infants' performance and that of adults (Best et al., 2001) is that whereas adults perceive these non-native distinctions in terms of native phonological contrasts, infants near the end of the first year do not yet pick up phonological information per se, having become attuned only to native phonetic-articulatory patterns, that is, native gesture constellations. Also consistent with the PAM/AO hypothesis, in Experiment 2 the older infants showed significantly lower levels of discrimination (i.e., a decline in discrimination) than the younger infants on both non-native and native within-organ contrasts in fricative voicing (isiZulu and English fricative voicing contrasts), but did not differ from the younger infants in discriminating a non-native between-organ contrast (Tigrinya place contrast). Given that the older infants showed a comparable decline, relative to younger infants, for native as well as non-native within-organ fricative voicing distinctions, but no decline for the non-native between-organ place distinction, it appears that between-versus within-organ differences may, in at least some cases, supersede the general effects of native language attunement.

Nonetheless, further research comparing the PAM/AO hypothesis to the predictions of the other models will be needed to confirm or refute this interpretation. Evidence about infants' perception of other types of within- and betwecn-organ distinctions would be especially informative. Declines in discrimination by 10–12 months have been found for a number of non-native consonant distinctions. For English-learning infants, this developmental pattern has already been found for Nthlakampx velar versus uvular ejectives, and for Hindi dental versus retroflex voiced stops and aspirated versus breathy-voiced dental stops, three contrasts that are also difficult for monolingual English-speaking adults (Werker et al., 1981; Werker & Lalonde, 1988; Werker & Tees, 1984, 1984a). In AO terms, each of these contrasts constitutes a within-organ minimal phonetic distinction: the Nthlakampx velar-uvular ejectives both involve stop closure of the tongue body near the back of the oral cavity (closed constriction, to use Articulatory Phonology terminology), the distinction being in its location (i.e., place of articulation: velar vs. uvular). The Hindi dental-retroflcx stops are instead distinguished by closed constrictions of the tongue tip that differ in location (dental vs. postalveolar), and the Hindi aspirated-breathy voiced stops are distinguished by different degrees of laryngeal constriction (i.e., fully vs. partially abducted vocal folds [wide vs. critical/narrow glottal constriction] for the aspirated and breathy voiced stops, respectively). Other stop voicing contrasts (i.e., within-organ laryngeal contrasts) appear to show a similar developmental decline. Computer-synthesized non-native VOT distinctions among stops (e.g., corresponding to wide vs. critical glottal constriction) appear to be discriminated by infants under six months who are learning English, Spanish, and Kikuyu (Aslin & Pisoni, 1980b; Lasky et al., 1975; Streeter, 1976; but see Eimas et al., 1971), but then show a decline in discrimination for English-learning infants (Eilers et al., 1979; but see Aslin & Pisoni, 1980a; Aslin & Pisoni, 1980b).

Various other within-organ non-native minimal contrasts likewise show language-related declines in discrimination during the second half-year. English-learning but not Spanish-learning infants display difficulty during the second half-year in discriminating the Spanish

alveolar tap versus trill contrast (Eilers, Gavin, & Oller, 1982; but see Jusczyk, Shea, & Aslin, 1984). The Spanish tap-trill contrast is a within-organ contrast, and is distinguished by differences in tongue tip stiffness during alveolar constriction (permitting a single [tap] versus multiple [trilled] contacts with the alveolar ridge). Japanese infants similarly show a developmental decline in discriminating the English liquids [ɹ]–[l], which are noncontrastive in their native language (Tsushima et al., 1994). These liquids display a within-organ distinction in tongue constrictions (for [ɹ]: dental/alveolar/retroflex approximation [narrow constriction] of tongue tip with narrow pharyngeal constriction of tongue body, Alwan, Narayanan, & Haker, 1997; Delattre & Freeman, 1968; for [l]: closed dental/alveolar constriction of the tongue tip with narrow uvular constriction of tongue body, Sproat & Fujimura, 1993). Finally, the English interdental fricative versus alveolar stop contrast has proven difficult for 6–8 and 10–12 month olds to discriminate, whether it is in their native language (English) or not (French); French — but not English-speaking adults also have difficulty with this contrast (Polka et al., 2001). It is a within-organ contrast involving a distinction in both location and degree of tongue tip constriction (respectively, critical [inter]dental constriction of the tongue tip [frication] vs. closed alveolar constriction of the tongue tip).

The converse, or between-organ, side of PAM/AO predictions are also consistent with findings of several non-native contrasts that 10–12 month olds discriminate *well.* The findings on the isiZulu dental versus lateral click contrast provide one such case. Infants discriminate this contrast well up to the oldest age tested, 14 months, as do English-speaking adults. Additionally, in the present report, infants up through 12 months discriminated the Tigrinya bilabial versus alveolar ejective stops. Both of those contrasts are between-organ distinctions. The isiZulu dental and lateral clicks are distinguished by tongue tip versus tongue body releases of the click's constrictions. (Both clicks have tongue tip and tongue body closures: Traill, 1985.) Tongue tip and body are different articulatory organs according to Articulatory Phonology; they can form constrictions independently, in the sense that one can engage in a constriction without the other necessarily constricting as well (Browman & Goldstein, in preparation; Browman & Goldstein, 1989, 1990; Hallé, 1982; Studdert-Kennedy & Goldstein, in press). The Tigrinya bilabial and alveolar ejectives arc minimally distinguished by closed constrictions of the lips (bilabial) versus tongue tip (alveolar), that is, a between-organ distinction. As for native contrasts, young infants discriminate the English labiodental versus interdental fricative contrasts [f]–[θ] and [v]–[ð] (Levitt, Jusczyk, & Carden 1988), which are acoustically similar and are fairly perceptually confusable by adults (Miller & Nicely, 1955). Both are between-organ distinctions involving critical constrictions [frication] of the lower lip versus the tongue tip, with the upper teeth as the target locus of constriction for each active articulator.

In sum, the re-examination of previous findings suggests converging support for the PAM/AO articulatory organ hypothesis. The pattern of results is compatible with the PAM assumption that native language experience has led to phonetic attunement, but not yet to true phonological knowledge, at 10–12 months. The other theoretical viewpoints summarized in Tables 1 and 2 appear to be less clearly supported by the developmental findings of the present report, in combination with the complete pattern of findings in the literature. However, further research will be needed to provide important additional tests of the PAM and articulatory organ hypotheses, using other types of non-native and native phonetic distinctions. For example, it is not clear whether within-organ minimal contrasts of constriction degree (e.g., stop vs. fricative at the same place of articulation) would show a 10–12 month decline in discrimination. However, Polka and colleagues' (2001) findings with English [d]–[ð], a within-organ contrast that confounds degree and location of constriction, suggest that at least some constriction degree contrasts, even native ones, may indeed pose difficulties for infants. PAM/AO should predict, to give another example, that the/b/-/v/labial manner/place contrast will show a decline in discrimination by infants learning a language such as Korean, which has/b/but lacks/v/.

Compatibly, anecdotal observations and evidence on English-to-Korean loanwords strongly suggest that Korean adults who are inexperienced with spoken English do indeed have difficulty with this within-organ contrast: they pronounce English/v/as [b] and appear to have difficulty discriminating English [b]–[v].

There are a few theoretical questions we would like to address briefly as we conclude this report. An obvious one is: Why should articulatory information play a more central role than purely acoustic information in infants' attunement to the language environment? Both PAM and the articulatory organ hypothesis posit that articulatory information is primary in infant speech perception. That premise calls for some clarification here. Although it is widely held that speech perception rests on the higher-level processing of acoustic information, we believe that a broader consideration of the ecological niche of human communicators instead points toward the detection of articulatory gestures, which have not only auditory but also visual and proprioceptive consequences, as the foundation for speech perception. By this reasoning, perception of articulatory information is more ecologically relevant, more basic, and more parsimonious than the processing of acoustic information per sc. In order to become native speakers, and not only listeners, of a language, infants must perceive which articulatory gestures native speakers are producing with their vocal tracts. This is surely a fundamental motivation for infants' responses to human speech, especially their attunement to the speech patterns of their language environment. Imitation and other forms of replication of adult utterances arc paramount to language learning (see also Studdert-Kennedy & Goldstein, 2003). The most parsimonious arrangement for the infant's task would be a very basic, direct link between speech perception and production. What better common currency could there be for this link than the articulatory gestures that directly shape the optic (and haptic, for felt speech) properties as well as the acoustic properties of a speech signal? The notion that infants can detect such articulatory commonalities across modalities is consistent with findings that young infants recognize relationships between visibly talking faces and audible speech signals (Kuhl & Meltzoff, 1982; Kuhl & Meltzoff, 1984; Rosenblum, Schmuckler, & Johnson, 1997; Walton & Bower, 1993).

A final question is whether there may be some sort of biological specialization that supports the detection of articulatory information in speech, and aids especially in infants' attunement to native speech. On the one hand, the PAM model has argued that the detection of articulatory gesture information in speech is not different in kind from the detection of distal information about other types of events (Best, 1995), a stance that is consistent with the direct realist assumptions of the ecological theory of perception (e.g., Gibson, 1979). Moreover, the articulatory organ hypothesis draws from Meltzoff's model of infants' imitation of facial movements. That is, imitation of speech is assumed to work on the same intermodal organ-matching skills as imitation of other facial actions (Meltzoff & Moore, 1997). On the other hand, however, much evidence supports a biological specialization of the left cerebral hemisphere for perception of specifically linguistic information in speech (e.g., Best & Avery, 1999). There is also evidence that vocal and facial imitation abilities may be unique, or developed to a substantially higher degree, in humans than in other primates (Hauser, 1996). Further research will be needed to identify the ways in which imitation, and neurobiological specialization for speech and language, may contribute to human infants' attunement to the native language near the end of the first year.

References

ALWAN AA, NARAYANAN SS, HAKER K. Toward articulalory-acoustic models for liquid consonants based on MRI and EPG data. Part II: The rhotics. Journal of the Acoustical Society of America 1997;101:1078–1089. [PubMed: 9035399]

ASLIN RN, PISONI DB. Effects of early linguistic experience on speech discrimination by infants: A critique of Eilers, Gavin, and Wilson (1979). Child Development 1980a;51:107–112. [PubMed: 7363730]

ASLIN, RN.; PISONI, DB. Some developmental processes in speech perception. In: Yeni-Komshian, GH.; Kavanaugh, JE.; Ferguson, CA., editors. Child Phonology Vol. 2: Perception. New York: Academic Press; 1980b. p. 67-96.

BENEDICT H. Early lexical development: Comprehension and production. Journal of Child Language 1979;6:183–200. [PubMed: 468932]

BEST, CT. Discovering messages in the medium: Speech and the prelinguistic infant. In: Fitzgerald, HE.; Lester, B.; Yogman, M., editors. Advances in pediatric psychology. New York: Plenum; 1984. p. 97-145.

BEST, CT. Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In: de Boysson-Bardics, B.; de Schonen, S.; Jusczyk, P.; MacNeilage, P.; Morton, J., editors. Developmental neurocognition: Speech and face processing in the first year of life. Dordrecht, The Netherlands: Academic Publishers B.V; 1993. p. 289-304.

BEST, CT. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In: Nusbaum, HC., editor. The development of speech perception: The transition from speech sounds to spoken words. Cambridge, MA: MIT Press; 1994a. p. 167-224.

BEST, CT. Learning to perceive the sound pattern of English. In: Rovee-Collier, C.; Lipsitt, L., editors. Advances in infancy research. Vol. 8. Hillsdale, NJ: Ablex Publishers; 1994b. p. 217-304.

BEST, CT. A direct realist perspective on cross-language speech perception. In: Strange, W.; Jenkins, JJ., editors. Cross-language speech perception. Timonium, MD: York Press; 1995. p. 171-204.

BEST CT, AVERY RA. Left hemisphere advantage for click consonants is determined by linguistic significance. Psychological Science 1999;10:65–69.

BEST CT, McROBERTS GW, GOODELL E. American listeners' perception of non-native consonant contrasts varying in perceptual assimilation to English phonology. Journal of the Acoustical Society of America 2001;109:775–794. [PubMed: 11248981]

BEST CT, McROBERTS GW, LAFLEUR R, SILVER-ISENSTADT J. Divergent developmental patterns for infants' perception of two non-native consonant contrasts. Infant Behavior and Development 1995;18:339–350.

BEST CT, McROBERTS GW, SITHOLE NM. Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants. Journal of Experimental Psychology: Human Perception and Performance 1988;14(3):345–360. [PubMed: 2971765]

BROWMAN CP, GOLDSTEIN L. Articulatory gestures as phonological units. Phonology 1989;6(2): 201–251.

BROWMAN CP, GOLDSTEIN L. Gestural specification using dynamically-defined articulatory structures. Journal of Phonetics 1990;18:299–320.

BROWMAN, C.; GOLDSTEIN, L. Articulatory phonology. Cambridge, MA: MIT Press; in preparation

BURNHAM DK. Developmental loss of speech perception: Exposure to and experience with a first language. Applied Psycholinguistics 1986;7:207–240.

CANONICI, NN. Imisindo Yesizulu: A simple introduction to Zulu phonology. Durban, South Africa: University of Natal; 1989.

CLEMENTS, N.; OSU, S. Explosives, implosives and nonexplosives: The linguistic function of air pressure differences in stops. In: Gussenhoven, C.; Warner, N., editors. Laboratory Phonology. Vol. 7. Berlin, Germany: Mouton de Gruyter; 2002.

COHEN, LB. An information-processing approach to infant perception and cognition. In: Simion, F.; Butterworth, G., editors. The development of sensory, motor and cognitive capacities in early infancy: From perception to cognition. Hove, England: Taylor & Francis; 1998. p. 277-300.

DELATTRE PC, FREEMAN D. A dialect study of American rs by X-ray motion picture. Linguistics 1968;44:29–68.

DIAMOND, A.; WERKER, JE.; LALONDE, C. Toward understanding commonalities in the development of object search, detour navigation, categorization, and speech perception. In: Dawson,

G.; Fischer, KW., editors. Human behavior and the developing brain. New York, NY: Guilford Press; 1994. p. 380-426.

DOKE CM. Phonetics of the Zulu language. Bantu Studies 1926;11:1–138.

EILERS RE. Context-sensitive perception of naturally produced stop and fricative consonants by infants. Journal of the Acoustical Society of America 1977;61:1321–1336. [PubMed: 889607]

EILERS RE, GAVIN WJ, OLLER DK. Cross linguistic perception in infancy: The role of linguistic experience. Journal of Child Language 1982;9:289–302. [PubMed: 7119036]

EILERS RE, GAVIN W, WILSON WR. Linguistic experience and phonemic perception in infancy: A cross-linguistic study. Child Development 1979;50:14–18. [PubMed: 446199]

EIMAS PD, SIQUELAND ER, JUSCZYK P, VIGORITO J. Speech perception in infants. Science 1971;171:303–306. [PubMed: 5538846]

FENNELL C, WERKER J. Early word learners' ability to access phonetic detail in well-known words. Language and Speech 2003;46:245–264. [PubMed: 14748446](this issue)

FERGUSON, CA.; MENN, L.; STOEL-GAMMON, C. Phonological development: Models, research, implications. Timonium, MD: York Press; 1992.

FOWLER CA. An event approach to the study of speech perception from a direct-realist perspective. Journal of Phonetics 1986;14:3–28.

GIANNINI A, PETTORINO M, TOSCANO M. Some remarks on the Zulu stops. Afrikanistische Arbeitspapiere 1988;13:95–116.

GIBSON, JJ. The ecological approach to visual perception. Boston: Houghton Mifflin; 1979.

GIBSON JJ, GIBSON EJ. Perceptual learning: Differentiation or enrichment? Psychological Review 1955;62(1):32–41. [PubMed: 14357525]

HALLÉ M. On distinctive features and their articulatory implementation. Natural Language and Linguistic Theory 1982;1:91–105.

HALLÉ PA, De BOYSSON-BARDIES B. Emergence of an early receptive lexicon: Infants' recognition of words. Infant Behavior and Development 1994;17:119–129.

HALLÉ PA, De BOYSSON-BARDIES B. The format of representation of recognized words in infants' early receptive lexicon. Infant Behavior and Development 1996;19:463–481.

HARNSBERGER JD. A cross-language study of the identification of non-native nasal consonants varying in place of articulation. Journal of the Acoustical Society of America 2000;108:764–783. [PubMed: 10955644]

HAUSER, MD. The evolution of communication. Cambridge, MA: MIT Press; 1996.

JUSCZYK, PW. Toward a model of the development of speech perception. In: Perkell, JS.; Klatt, DH., editors. Invariance and variability in speech processes. Hillsdale, NJ: Lawrence Earlbaum Associates, Publishers; 1986. p. 1-35.

JUSCZYK PW. From general to language-specific capacities: The WRAPSA Model of how speech perception develops. Journal of Phonetics 1993;21:3–28.

JUSCZYK, PW. Infant speech perception and the development of the mental lexicon. In: Goodman, JC.; Nusbaum, HC., editors. The development of speech perception: The transition from speech sounds to spoken words. Cambridge, MA: MIT Press; 1994. p. 227-270.

JUSCZYK, PW. The discovery of spoken language. Cambridge, MA: Bradford Books: MIT Press; 1997.

JUSCZYK PW, BERTONCINI J. Viewing the development of speech perception as an innately guided learning process. Language and Speech 1988;31(3):217 – 238. [PubMed: 3078979]

JUSCZYK PW, LUCE PA, CHARLES-LUCE J. Infants' sensitivity to phonotactic patterns in the native language. Jounal of Memory and Language 1994;33:630–645.

JUSCZYK PW, SHEA SL, ASLIN RN. Linguistic experience and infant speech perception: A re-examination of Eilers, Gavin, and Oller (1982). Journal of Child Language 1984;2:453 – 466. [PubMed: 6746786]

KUHL, PK. Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In: de Boysson-Bardies, B.; de Schonen, S.; Jusczyk, P.; McNeilage, P.; Morton, J., editors. Developmental neurocognition: Speech and face processing in the first year of life. Dordrecht: Kluwer Academic Publishers; 1993. p. 259-274.

KUHL PK, MELTZOFF AN. The bimodal perception of speech in infancy. Science 1982;218 (December):1138–1141. [PubMed: 7146899]

KUHL PK, MELTZOFF AN. The intermodal representation of speech in infants. Infant Behavior and Development 1984;7:361–381.

KUHL PK, WILLIAMS KA, LACERDA F, STEVENS KN, LINDBLOM B. Linguistic experience alters phonetic perception in infants by six months of age. Science 1992;255(January):606–608. [PubMed: 1736364]

LADEFOGED, P.; MADDIESON, I. The sounds of the world's languages. Oxford: Blackwell; 1996.

LALONDE CE, WERKER JF. Cognitive influences on cross-language speech perception in infancy. Infant Behavior & Development 1995;18:459 – 475.

LASKY RE, SYRDAL-LASKY A, KLEIN RE. VOT discrimination by four to six and a half month old infants from Spanish environments. Journal of Experimental Child Psychology 1975;20:215–225. [PubMed: 1194843]

LEVITT A, JUSCZYK PW, CARDEN G. Context effects in two-month-old infants' perception of labiodental/interdental fricative contrasts. Journal of Experimental Psychology 1988;14:361–368. [PubMed: 2971766]

MACKAIN K. Assessing the role of experience on infants' speech discrimination. Journal of Child Language 1982;9:527–542. [PubMed: 7174755]

MADDIESON, I. Patterns of sounds. New York: Cambridge University Press; 1984.

MAYE J, WERKER JF, GERKEN L. Infant sensitivity to distributional information can affect phonetic discrimination. Cognition 2002;82:B101 – B111. [PubMed: 11747867]

MELTZOFF A, MOORE M. Explaining facial imitation: A theoretical model. Early Development and Parenting 1997;6:179 – 192.

MILLER C. Developmental changes in male/female voice classification by infants. Infant Behavior and Development 1983;6:313 – 330.

MILLER GA, NICELY PE. Perceptual confusions among some English consonants. Journal of the Acoustical Society of America 1955;27:338 – 352.

PEGG JE, WERKER JF. Adult and infant perception of two English phones. Journal of the Acoustical Society of America 1997;102:3742–3753. [PubMed: 9407666]

PEPERKAMP, S.; DUPOUX, E. Coping with phonological variation in early lexical acquisition. In: Lasser, I., editor. The process of language acquisition. Berlin, Germany: Peter Lang, Verlag; 2002. p. 359-385.

PIRELLO K, BLUMSTEIN S, KUROWSKI K. The characteristics of voicing in syllable-initial fricatives in American English. Journal of the Acoustical Society of America 1997;101:3754–3765w. [PubMed: 9193062]

POLKA L. Characterizing the influence of native experience on adult speech perception. Perception and Psychophysics 1992;52:37–52. [PubMed: 1635856]

POLKA L, COLANTONIO C, SUNDARA M. A cross-language comparison of /d/-/ð/ perception: Evidence for a new developmental pattern. Journal of the Acoustical Society of America 2001;109:2190–2201. [PubMed: 11386570]

POULOS, G.; BOSCH. Zulu. München, Germany: Loncom Europa; 1997.

ROSENBLUM LD, SCHMUCKLER MA, JOHNSON JA. The McGurk effect in infants. Perception and Psychophysics 1997;59:347–357. [PubMed: 9136265]

ROUX, JC. On ingressive glottalic and velaric articualtions in Xhosa. Paper presented at the International Conference of Phonetic Sciences; Aix-en-Provence, France. 1991.

SLIS IH, COHEN A. On the complex regulating the voiced-voiceless distinction. Language & Speech 1969;12:80 – 102. [PubMed: 5792375]

SPROAT R, FUJIMURA O. Allophonic variation in English /l/ and its implications for phonetic implementation. Journal of Phonetics 1993;21:291 – 311.

STAGER CL, WERKER JF. Infants listen for more phonetic detail in speech perception than in word-learning tasks. Nature 1997;388:381 – 382. [PubMed: 9237755]

STREETER L. Language perception of two-month-old infants shows effects of both innate mechanisms and experience. Nature 1976;259:39–41. [PubMed: 1256541]

STUDDERT-KENNEDY, M. How did language go discrete?; Paper presented at the Fourth International Conference on the Evolution of Language; Harvard University, Cambridge, MA. 2002.

STUDDERT-KENNEDY, M.; GOLDSTEIN, L. Lanuching language: The gestural origin of discrete infinity. In: Christiansen, M.; Kirby, S., editors. Language evolution: The states of the art. Oxford U.K: Oxford University Press; in press

SWINGLEY D. Phonetic detail in the developing lexicon. Language and Speech 2003;46:265–294. [PubMed: 14748447](this issue)

SWINGLEY D, ASLIN RN. Lexical neighborhoods in the word-form representations of 14-month-olds. Psychological Science 2002;13:480–484. [PubMed: 12219818]

TEES RC, WERKER JF. Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. Canadian Journal of Psychology 1984;38:579–590. [PubMed: 6518419]

TRAILL, A. Phonetic and phonological Studies of !Xóõ Bushman. Hamburg: Helmut Buske Verlag; 1985.

TRAILL A, KHUMALO JSM, FRIDJHON P. Depressing facts about Zulu. African Studies 1987;46:255–274.

TREHUB SE. The discrimination of foreign speech contrasts by infants and adults. Child Development 1976;47(1):466 – 472.

TSAO, F.; LIU, H.; KUHL, PK.; TSENG, C. *Perceptual discrimination of Mandarin Fricative-affricate contrast by English-learning and Mandarin-learning infants*. Paper presented at the International Conference on Infant Studies; Brighton, U.K. 2000.

TSUSHIMA, T.; TAKIZAWA, O.; SASAKI, M.; SHIRAKI, S.; NISHI, K.; KOHNO, M.; MENYUK, P.; BEST, CT. *Discrimination of English /r-1/and /w-j/ by Japanese infants at 6–12 months: Language-specific developmental changes in speech perception abilities*. Paper presented at the International Conference on Spoken Language Processing; 1994.

Van WYCK, EB. Praktiese fonetik vir taalstudente: 'n inleiding (Practical phonetics for students: An introduction). Durban, South Africa: Butterworth; 1979.

WALTON GE, BOWER TGR. Amodal representation of speech in infants. Infant Behavior and Development 1993;16:233–243.

WERKER JF. Becoming a native listener. Am Sci 1989;77:54–59.

WERKER JF, GILBERT JHV, HUMPHREY K, TEES RC. Developmental aspects of cross-language speech perception. Child Development 1981;52:349–355. [PubMed: 7238150]

WERKER JF, LALONDE CE. Cross-language speech perception: Initial capabilities and developmental change. Develop Psych 1988;24(5):672–683.

WERKER, JF.; PEGG, JE. Infant speech perception and phonological aquisition. In: Ferguson, CA.; Menn, L.; Stoel-Gammon, C., editors. Phonological development: Models, research, implications. Timonium, MD: York Press; 1992. p. 285-311.

WERKER JF, TEES RC. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behavior and Development 1984a;7:49 – 63.

WERKER JF, TEES RC. Phonemic and phonetic factors in adult cross-language speech perception. Journal of the Acoustical Society of America 1984b;75:1866 – 1878. [PubMed: 6747097]

YOUNGER BA, COHEN LB. Infant perception of correlations among attributes. Child Development 1983;54:858–869. [PubMed: 6617307]

ZIERVOGEL, D.; LOUW, JA.; TALJAARD, PC. A handbook of the Zulu language. Johannesburg, South Africa: J. L. van Schaik; 1985.

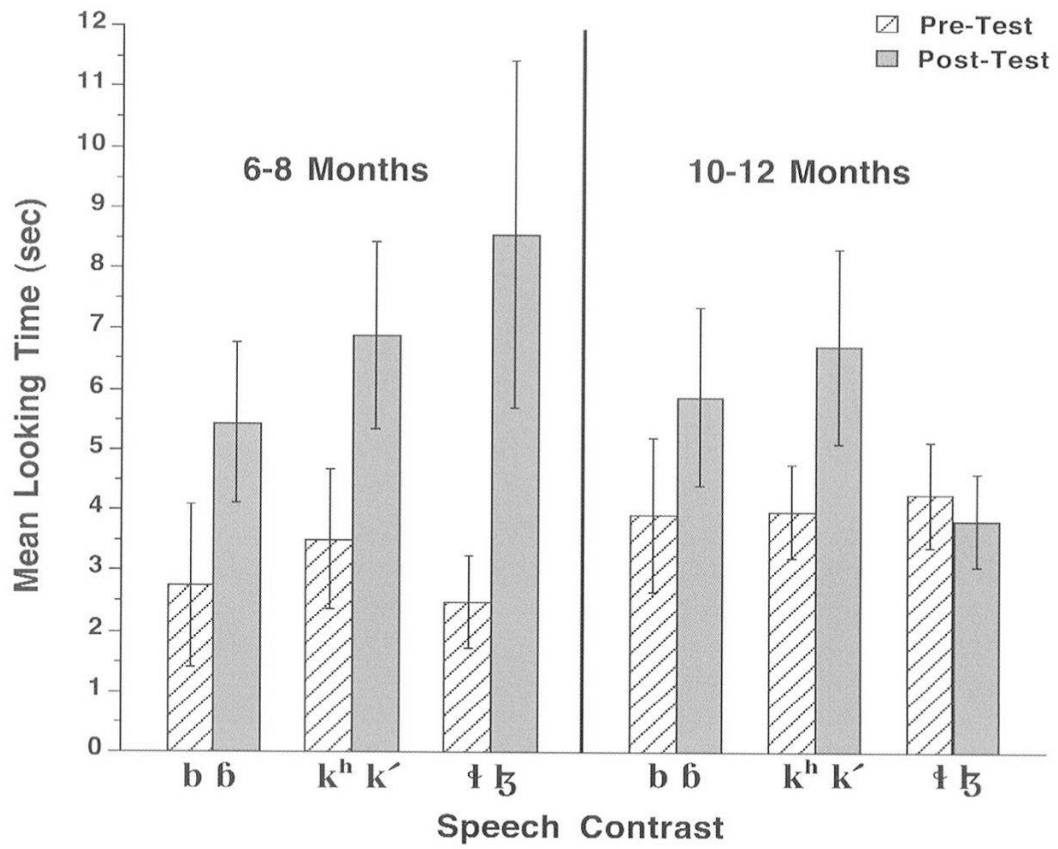**Figure 1.**
Discrimination of the three isiZulu consonant contrasts by 6–8 and 10–12 month old American infants, Experiment 1
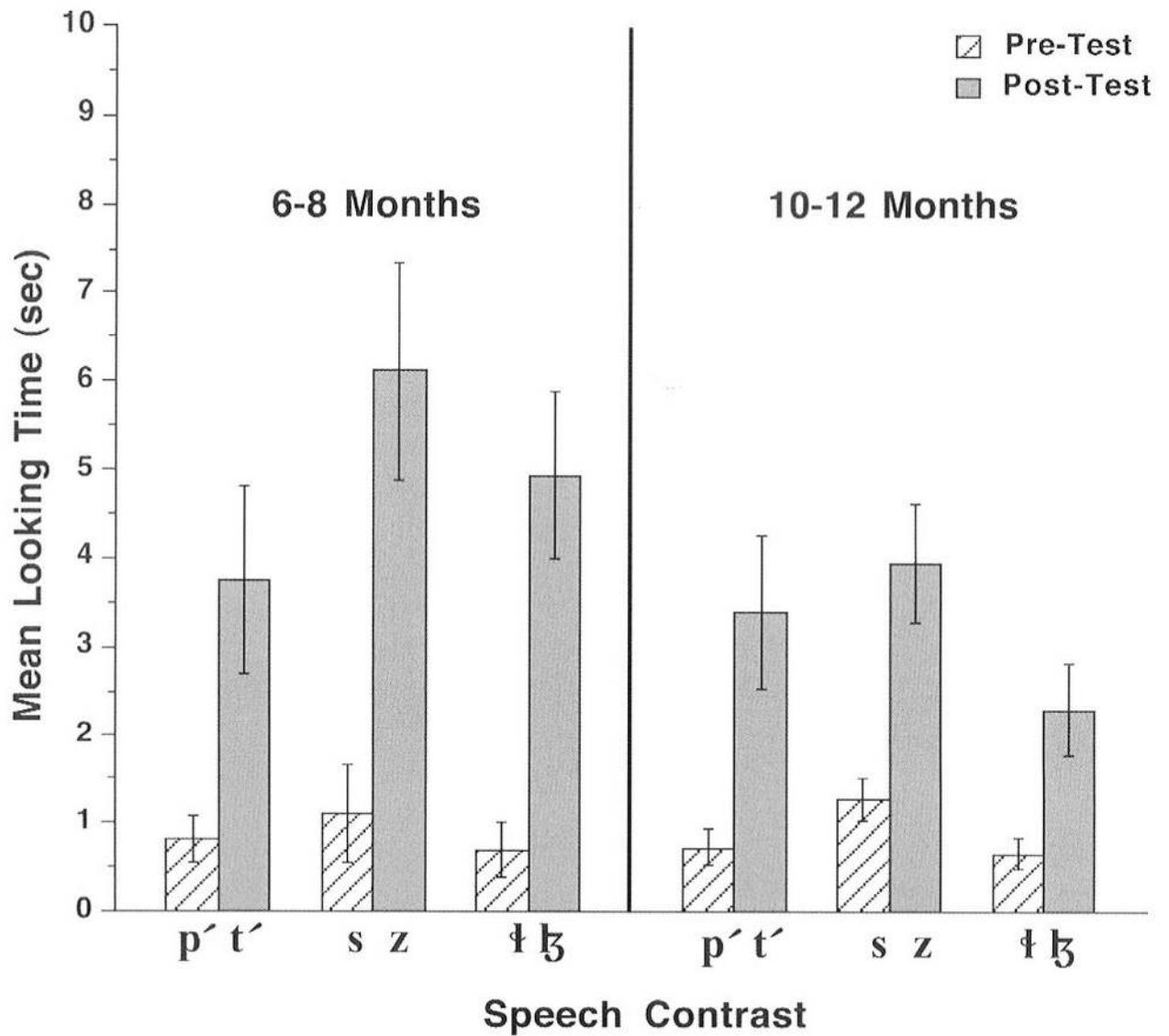
Figure 2.
Discrimination of the English, Tigrinya and isiZulu stimulus contrasts by 6–8 and 10–12 month old American infants, Experiment 2

TABLE 1

Predictions from various theoretical perspectives regarding discrimination of the three isiZulu within-organ consonant contrasts at 10–12 months (Experiment l), by comparison to 6–8 months

Hypotheses:	Discrimination At 10–12 months:		
	Bilabial stops	**Velar stops**	**Lateral fricatives**
General mechanism accounts:			
Fragile-Robust (Burnham)	very good	very good	very good
Auditory Tuning (Kuhl)	failure	some decline	very good
General Cognitive (Werker et al.; Cohen et al.)	failure	very good	failure
Linguistic accounts:			
Phonological/PAM adult (Jusczyk; Pegg/Werker; Best)	failure	some decline	very good
Allophonic Experience (Tees/Werker)	very good	some decline	failure
Native Word Phonetics (Stager/Werker)	failure	some decline	failure
PAM/Articulatory Organ (Best/Goldstein)	decline	decline	decline

TABLE 2

Critical properties of the consonant contrasts used in Experiment 2, and predictions of various models

	Tigrinya ejectives: place of articulation	English fricatives: voicing	isiZulu fricatives: voicing
Properties of contrasts:			
Language experience:	non-native	native	non-native
Articulatory organ distinction:	between-organ (lips vs. tongue tip)	within-organ (laryngeal)	within-organ (laryngeal)
Predictions:			
phonotactic learning (Jusczyk et al.)	decline at 10–12 months	decline at 10–12 months	decline at 10–12 months
fricative voicing (Eilers)	(irrelevant to hypoth.)	failure at both ages	failure at both ages
PAM/AO (Best/Goldstein)	very good at both ages	decline at 10–12 months	decline at 10–12 months