Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system

Catherine T. Best[a],
Department of Psychology, Wesleyan University, Middletown, Connecticut 06459 and Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511

Gerald W. McRoberts, and
Department of Psychology, Lehigh University, Bethlehem, Pennsylvania 18015 and Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511

Elizabeth Goodell[b]
Department of Psychology, University of Connecticut, Storrs, Connecticut 06269 and Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511

Abstract

Classic non-native speech perception findings suggested that adults have difficulty discriminating segmental distinctions that are not employed contrastively in their own language. However, recent reports indicate a gradient of performance across non-native contrasts, ranging from near-chance to near-ceiling. Current theoretical models argue that such variations reflect systematic effects of experience with phonetic properties of native speech. The present research addressed predictions from Best's perceptual assimilation model (PAM), which incorporates both contrastive phonological and noncontrastive phonetic influences from the native language in its predictions about discrimination levels for diverse types of non-native contrasts. We evaluated the PAM hypotheses that discrimination of a non-native contrast should be near-ceiling if perceived as phonologically equivalent to a native contrast, lower though still quite good if perceived as a phonetic distinction between good versus poor exemplars of a single native consonant, and much lower if both non-native segments are phonetically equivalent in goodness of fit to a single native consonant. Two experiments assessed native English speakers' perception of Zulu and Tigrinya contrasts expected to fit those criteria. Findings supported the PAM predictions, and provided evidence for some perceptual differentiation of phonological, phonetic, and nonlinguistic information in perception of non-native speech. Theoretical implications for non-native speech perception are discussed, and suggestions are made for further research.

I. INTRODUCTION

Adults' perception of speech contrasts is strongly influenced by experience with the phonological system of their native language (e.g., Abramson and Lisker, 1970). A traditional account for this phenomenon has been a perceptual version of the concept that a native-language "phonological filter" operates in production of non-native segments (Polivanov, 1931; Trubetskoy, 1939/1969). That is, it has been assumed that mature listeners have difficulty discriminating phonetic distinctions that do not occur as a native phonological contrast. Perhaps

acbest@wesleyan.edu.
[b]Present address: Landmark School, 429 Hale St., Prides Crossing, MA 01965.

the most widely cited example of such perceptual difficulty is the poor discrimination of English /r/-/l/ by speakers of languages that lack this contrast, such as Japanese (e.g., Goto, 1971; Miyawaki *et al.*, 1975; Mochizuki, 1981; Best and Strange, 1992; MacKain *et al.*, 1981). Similarly poor non-native speech perception performance has been documented for speakers of other languages. For example, English speakers have difficulty discriminating contrasts such as Hindi retroflex versus dental stops and Nthlakampx velar versus uvular ejectives (Werker *et al.*, 1981; Werker and Tees, 1984). Discrimination in such cases is near chance, in striking contrast to the ceiling-level performance typically found with native language distinctions.

What is it about native language experience that results in such difficulties with non-native speech discrimination? One type of explanation emphasized exposure in early development as being critical to the "tuning" of relevant sensori-neural mechanisms. For example, some argued that innate, linguistically specialized neural mechanisms, initially tuned to universal settings of phonetic categories and/or boundaries, are modified by early exposure to specific phonetic features (e.g., Eimas, 1975, 1991). Others posited the nonlinguistic view that early exposure to specific acoustic properties maintains or enhances the tuning of prewired psychophysical mechanisms that respond selectively to those properties (e.g., Aslin and Pisoni, 1980). Such prewired mechanisms are generally assumed, by the latter view, to be components of general auditory processing skills that are part of our mammalian (or vertebrate) evolutionary endowment (e.g., Kuhl, 1988; Dooling, 1989).

It has since become apparent, however, that neither account of critical early tuning can adequately explain all aspects of adults' non-native speech perception. Numerous studies have shown that discrimination of unfamiliar phonetic contrasts can be improved even in adults through extensive natural experience, intensive laboratory training, or experimental manipulations that reduce task memory demands (e.g., Logan *et al.*, 1991; Lively *et al.*, 1993; Pisoni *et al.*, 1982; MacKain *et al.*, 1981; Strange and Dittman, 1984; Werker and Logan, 1985; Werker and Tees, 1984). That is, exposure need not occur early in development—even limited exposure in adulthood can improve performance to some extent.[1] In response to those findings, some have proposed that language experience affects higher-level processes that remain malleable, such as phonological encoding or memory retention, rather than lower-level sensorineural responsivities that are relatively permanently changed by early experience (Tees and Werker, 1984; Werker and Tees, 1984).

Further damaging to the sensorineural tuning hypotheses is the fact that early exposure to specific phonetic or acoustic properties, or lack thereof, does not guarantee good versus poor discrimination, respectively. For example, discrimination of non-native stop voicing contrasts is poor in American English listeners (Abramson and Lisker, 1970), despite the fact that the range of voice onset times (VOTs) involved is amply manifested in English stop allophones (cf. MacKain, 1982), and adults' discrimination levels for non-native contrasts are not systematically related to whether or not the associated phonetic features occur within the listeners' native language (Polka, 1992). Native English speakers' discrimination of four Hindi dental-retroflex stop consonant contrasts differing in voicing (prevoiced, voiceless unaspirated, voiceless aspirated, and breathy voiced) varied between poor and excellent, irrespective of whether the specific voicing type occurs in English (Polka, 1991). Conversely, adults discriminate certain non-native contrasts quite well, even with virtually no prior exposure to their distinctive phonetic-acoustic features in speech. American English listeners show very

---

[1]Still, it is important to note that laboratory training effects are limited in magnitude in adults (Lively *et al.*, 1994), and that listeners' discrimination of the critical acoustic properties of nonnative contrasts presented in isolation may fail to generalize to good discrimination of them within speech contexts (Miyawaki *et al.*, 1975; Werker and Tees, 1984).

good to excellent discrimination of Zulu click consonants, despite their lack of experience or training with such clicks (Best *et al.*, 1988).

Two important conclusions can be drawn from these findings: Adult discrimination of non-native speech contrasts is *not* uniformly poor, as the classic view would have it. Instead, we see wide variation in performance level, from poor to excellent, that does not depend on the presence or absence of the critical phonetic/acoustic features in native speech.

If early sensorineural tuning fails to explain the variation in non-native speech discrimination, then what does account for it? Several recent theoretical models posit that native speech experience provides an organizing perceptual framework that shapes discrimination of unfamiliar speech contrasts. Best's perceptual assimilation model (PAM: Best, 1994a, b, 1995; Best *et al.*, 1988), Flege's speech learning model (SLM: Flege, 1986, 1990, 1995; cf. Guion *et al.*, 2000), and Kuhl's Native Language Magnet model (NLM: Grieser and Kuhl, 1989; Iverson and Kuhl, 1996; Kuhl, 1991, 1992; Kuhl *et al.*, 1992) all presume that adults' discrimination of non-native speech contrasts is systematically related to their having acquired a native speech system. However, the models differ in how they conceive of the native perceptual framework, as in ongoing theoretical debates about speech perception. One basic view is that speech perception depends on the same general-purpose auditory processes employed for perception of nonspeech sounds (e.g., Diehl and Kluender, 1989; Kluender, 1994; Kuhl, 1988; cf. theorical overview in Best, 1995). An opposing view is that a specialized linguistic-phonetic module is involved in perception of speech alone (Liberman *et al.*, 1967; Liberman and Mattingly, 1989). Also under debate is whether the perceptual mechanisms, whether general or specialized, operate on acoustic (e.g., Diehl and Kluender, 1989; Kuhl, 1988, 1991, 1992) or articulatory information (e.g., Best, 1995; Fowler, 1986, 1989; Liberman and Mattingly, 1989).

SLM addresses primarily how adult speakers acquire phonological segments for a second language (L2), particularly in production and particularly by relatively experienced L2 speakers. It proposes that non-native phones are "equivalence-classified" relative to native language (L1) phonemes on the basis of phonetic similarity. New L2 phonological categories are more likely to be developed, hence produced (and perceived) fairly accurately, the more dissimilar they are from the closest native phonemes. SLM remains neutral regarding general versus specialized mechanisms and extraction of acoustic versus linguistic-phonetic information from speech.

NLM instead proposes that early in life, listeners develop acoustic prototypes for native phonemic categories. NLM assumes that speech perception involves general auditory mechanisms that process acoustic rather than specifically phonetic information. In NLM, native prototypes have magnetlike effects, in which the nearby perceptual space is "shrunk," making it more difficult to discriminate phonetic variation around prototypes than around non-prototypes, or poor exemplars, of the same category. So NLM, unlike SLM, predicts an asymmetry for discriminating prototypical versus non-prototypical stimuli. Listeners fail to develop prototypes for non-native categories, due to lack of relevant acoustic experience. Hence within-category discrimination for non-native phones is expected to be uniform rather than asymmetrical.

Both SLM and NLM have contributed importantly to our understanding of the perception of non-native speech segments, and have generated substantial research. In a quest to account for discrimination of nonnative *contrasts*, however, they both have limitations. SLM's primary shortfall is its focus on individual phonemes—it makes no explicit predictions about discrimination of non-native contrasts (Flege, personal communication). As for NLM, recent concerns have been raised that the perceptual magnet effect may not be robust across listener

groups (Lively, 1993; cf. Frieda *et al.*, 1999). Moreover, given listeners' frequent identification of the "non-prototype" stimuli as exemplars of a different category altogether than the prototype, the asymmetry in discrimination may actually reflect better between-category than within-category discrimination (Lotto *et al.*, 1996; Sussman and Lauckner-Morano, 1995), that is, the classic phenomenon of categorical perception (Lotto *et al.*, 1998). Additionally, NLM espouses basic principles of the early tuning accounts which, as we summarized earlier, are problematic. Neither SLM nor NLM comprehensively explains the variations in non-native discrimination that cannot be traced to the presence versus absence of features in the listeners' language, as discussed earlier.

Those theoretical gaps are addressed by the perceptual assimilation model (PAM). That model was originally developed to account for the previously unexpected finding that American English listeners discriminate Zulu clicks quite well; the authors hypothesized that this was due to the fact that they had perceived the clicks as nonspeech sounds (Best *et al.*, 1988). Of the three non-native speech perception models, only PAM makes explicit predictions about assimilation and discrimination differences for diverse types of non-native contrasts. And PAM alone incorporates principles of phonological theory, the branch of linguistics that concerns the linguistic function and structure of the native system of phonological contrasts. The specific phonological theory that PAM draws from is articulatory phonology (Browman and Goldstein, 1986, 1989, 1990a, b, 1992), which is compatible with PAM's direct realist (ecological) position that what listeners detect in speech is information regarding the articulatory gestures that generated the signal (e.g., Best, 1995; Fowler, 1986, 1989; Fowler *et al.*, 1990). Gestures are defined by the articulatory organs (active articulator, including laryngeal gestures), constriction locations (place of articulation), and constriction degree (manner of articulation) employed. We will use the term "native phoneme" to refer to a functional equivalence class of articulatory variants that serve a common phonological function, as evidenced by their contribution to distinguishing lexical items, identifying morphosyntactic units, and participating in other phonological alternations such as context-conditioned allophony. PAM posits that non-native speech perception is strongly affected by listeners' knowledge (whether implicit or explicit) of native phonological equivalence classes, and that listeners perceptually assimilate non-native phones to native phonemes whenever possible, based on detection of commonalities in the articulators, constriction locations and/or constriction degrees used (Best, 1993, 1994a, b, 1995).

According to PAM (see Best, 1995), a given non-native phone may be perceptually assimilated to the native system of phonemes in one of three ways: (1) as a **C**ategorized exemplar of some native phoneme, for which its goodness of fit may range from excellent to poor; (2) as an **U**ncategorized consonant or vowel that falls somewhere in between native phonemes (i.e., is roughly similar to two or more phonemes); or (3) as a **N**onassimilable nonspeech sound that bears no detectable similarity to any native phonemes. Adults' discrimination of a non-native contrast is predicted to depend on how each of the contrasting phones is assimilated. Several pairwise assimilation types are possible. The non-native phones may be phonetically similar to two different native phonemes and assimilate separately to them, which was termed **T**wo **C**ategory assimilation[2] (**TC**). Both may, instead, assimilate equally well or poorly to a single native phoneme, termed **S**ingle **C**ategory assimilation (**SC**). Or both might assimilate to a single native phoneme, but one may fit better than the other, termed a **C**ategory **G**oodness difference (**CG**). Alternatively, one non-native phone may be **U**ncategorized, as defined above, while the

---

[2]Terminology for PAM predictions originally referred to "phonetic categories." However, given the model's ecological theoretical perspective, it does not espouse cognitive processing assumptions about mental representations of categories, category formation, etc. Therefore, in this article we have discussed native phonological influences in terms of functional equivalence classes rather than in terms of phonetic categories. However, we have retained the assimilation terms used in earlier presentations of PAM, for consistency with previous publications.

other is **C**ategorized, forming an **U**ncategorized–**C**ategorized pair (**UC**). Or both non-native phones might be **U**ncategorized speech segments (**UU**). Finally, the two phones' articulatory properties may both be quite discrepant from any native phonemes, and be perceived as **N**on-**A**ssimilable (**NA**) nonspeech sounds.

As for discrimination of non-native contrasts, it can be hindered, aided, or unaffected by native phonology, depending on how the non-native phones relate to native phonemes and contrasts (Best, 1994a, 1995). Native phonology should aid discrimination when the two phones are separated by native phonological boundaries, but should hinder it when both phones assimilate to the same native phoneme. However, discrimination of non-native elements that are heard as nonspeech sounds is neither helped nor hindered by native phonology. NA contrasts, therefore, are predicted to show good to excellent discrimination, depending on their perceived differences as nonspeech sounds. However, TC and UC contrasts should be discriminated quite well because in both cases the contrasting phones fall on opposite sides of a native phonological boundary. On the other hand, with CG and SC types, both phones assimilate to the same native phoneme, so discriminability is hindered by native phonology. If one phone is a good fit and the other is poor, discrimination will be very good (CG difference), but not as good as in TC contrasts because it is hindered by assimilating to a single native phoneme. In SC cases, both non-native phones are equivalent in phonetic goodness, hence discrimination is poor, hindered both by lack of phonological contrast and by lack of difference in fit. For example, Japanese speakers are likely to assimilate English /r/ and /l/ as poor examples of a single Japanese phoneme (/r/ or perhaps /w/: Best and Strange, 1992; Takagi and Mann, 1995; Yamada and Tohkura, 1992), and discriminate the /r/-/l/ contrast poorly. For uncategorized-uncategorized (UU) assimilations, discrimination is less strongly affected by native phonological equivalence classes, and should range between fair and good, dependent on perceived similarity of the non-native phones to each other and to the set of nearby native phonemes.

The current research focuses on those contrasts involving only non-native phones that are perceptually categorized to native phonemes, as defined above, that is, the TC, CG, and SC assimilation types. PAM predicts the following gradient of discrimination levels for these: TC>CG>SC (Best, 1994a, 1995). PAM's predictions about each of these assimilation types have been supported by a number of cross-language perception studies (see Best, 1994a, b, 1995). As noted, English speaking adults fail to assimilate Zulu click consonants to English consonants, instead perceiving them as nonspeech sounds, consistent with the NA pattern. In keeping with PAM's predictions about non-native NA contrasts, discrimination of the clicks is good to very good (Best *et al.*, 1988). English listeners' perception of clicks as non-speech is supported by recent evidence that whereas Zulu listeners show right ear superiority for click discrimination in a dichotic listening task, presumed to reflect left hemisphere language specialization, American English listeners do not (Best and Avery, 1999). Moreover, English-learning infants fail to show a developmental decline in discrimination of the clicks by 10–12 months (Best *et al.*, 1988) comparable to that found for other non-native consonant contrasts (e.g., Werker, 1989; Werker and Pegg, 1992; Werker *et al.*, 1981; Werker and Lalonde, 1988). In particular, 10–12 month olds discriminated a click contrast but failed to do so with a non-native contrast from Werker *et al.* (1981) on which adults' perception had been consistent with SC assimilation (Best *et al.*, 1995). In cross-language studies of adults' non-native speech perception, Japanese listeners displayed SC assimilation of American English /r/-/l/ and CG assimilation of English /w/-/r/, with better discrimination of the latter, as predicted by PAM (Best and Strange, 1992). French listeners categorized and discriminated English /w/-/r/ in a CG pattern, consistent with French and English /r/ articulatory differences (Hallé *et al.*, 1999).

Studies from other research groups also are consistent with certain PAM predictions. In her study of English listeners' perception of four Hindi dental-retroflex stop contrasts differing in

voicing type, Polka (1991) reported that, based on listeners' descriptions of the contrasts, SC-type assimilations were associated with lower discrimination performance than TC-type assimilations, as PAM predicts. She also found that English listeners tended to assimilate Farsi voiced velar versus uvular stops (/g/-/G/) as a CG contrast and Salish velar versus uvular ejectives (/k′/-/q′/) as a SC (or NA[3]) contrast, with a tendency toward better discrimination of the former distinction, as would be expected according to PAM (Polka, 1992)[).4] In another recent study, Japanese listeners' discrimination of UU and UC assimilations of English consonant contrasts fit PAM predictions in all but one UC case (Guion *et al.*, 2000). Interestingly, two studies of early bilinguals revealed poor discrimination of contrasts that fit a SC pattern with respect to the L1, but a TC pattern with respect to the L2, suggesting long-term effects of L1 phonological organization even in listeners who have been fluent in the L2 from a young age (Calderón and Best, 1996; Pallier *et al.*, 1997).

However, no findings have yet been published on non-native contrasts that clearly fit the TC pattern, i.e., in which neither non-native phone is a good match for a native phoneme yet the both are perceptually assimilated to two different native phonological classes. Evidence on the TC pattern is important, given that the predicted excellent, nativelike (or nearly so) levels of discrimination and categorization performance would be quite unexpected according to the more traditional assumption that adults should have difficulty discriminating any contrasts that do not occur in the native language (Polivanov, 1931). Relatedly, reports are still lacking on systematic comparison of TC, CG, and SC assimilation types needed to evaluate PAM's strong prediction for significantly better discrimination of TC than CG assimilation types, which in turn should show better discrimination of SC types. Alternative outcome patterns remain possible for those three types of non-native contrast. One is that there could be equally poor discrimination for the three types of contrast, as suggested by traditional claims about non-native speech perception. As summarized earlier, this outcome is highly unlikely in light of previous findings that discrimination levels can differ substantially among non-native contrasts. Another possibility might be that discrimination differences could be determined by some other factor, such as acoustic differences among the contrasts, and not by their phonological assimilations. To evaluate these possibilities, Experiment 1 systematically compared discrimination levels among non-native contrasts that were expected to yield TC, CG, and SC assimilation patterns.

II. EXPERIMENT 1

To optimize comparisons of performance among SC, CG, and TC assimilations, all three stimulus contrasts were taken from a single language, Zulu. None were phonological contrasts in English. All three were differentiated by laryngeal gestures. The goal was to include one non-native contrast that American English (AE) listeners were likely to assimilate to two contrasting English phonemes (TC), another that they should assimilate as a noticeable category goodness difference within a single English phoneme (CG), and a third that they should assimilate with nearly equal fit to a single phoneme in English (SC). The following contrasts were selected, based on their articulatory-phonetic characteristics relative to English (Ladefoged and Maddieson, 1996; Maddieson, 1984; Ruhlen, 1975):

   i.   voiceless versus voiced lateral fricatives (/ɬ/-/ɮ/);

   ii.   voiceless aspirated versus ejective (glottalized) velar stops (/kʰ/-/k′/);

   iii. plosive versus implosive voiced bilabial stops (/b/-/ɓ/).

---

[3]Based on Polka's summary of these subjects' descriptions, the latter assimilations may actually have been UU rather than NA types, given that they reported hearing consonants and/or vowels rather than nonspeech sounds.
[4]It should be noted, however, that Polka concluded that certain other aspects of her findings in these two studies may have been guided by acoustic attributes of the stimuli rather than by phonological/phonetic properties of the listeners' native language.

The Zulu lateral fricative contrast uses a place of articulation that is non-native for AE fricatives, though the articulatory organs and constriction locations involved are similar to AE / l/. Voiceless–voiced fricative distinctions do occur in AE at other constriction locations. In both languages, fricative voicing contrasts are signaled by a laryngeal gesture of glottal abduction (voiceless), versus a glottal setting that results in vocal fold vibration (voiced), during frication at the supralaryngeal constriction location. In the other two Zulu contrasts, the laryngeal distinction itself rather than the constriction location was non-native. Location for the Zulu velar stop constriction corresponds to that for AE /k/. The glottal abduction for Zulu / k/ makes it essentially identical to AE /k/; both are narrowly transcribed as [k^h], i.e., long-lag voice onset with positive airflow through the open glottis during release of velar closure, resulting in aspiration. The distinctive laryngeal gesture for the contrasting ejective /k′/ is a glottal adduction, with a resulting (near-)cessation of glottal airflow during release of the velar stop closure. The latter laryngeal gesture is not used in utterance-initial AE stops (although some speakers produce ejectives in forceful releases of utterance-final voiceless stops). As for the third Zulu contrast, the glottal setting is similar in Zulu and AE /b/, in that Zulu /b/ displays a short unaspirated voicing lag (i.e., [p]), as is the case for the common [p] allophone of AE / b/ (which can also be realized as fully voiced [b]). The implosive Zulu /ɓ/ involves voicing during bilabial closure and release (as in the AE [b] allophone), but adds a simultaneous rapid lowering of the larynx, which causes a brief negative airflow during release. Larynx lowering is not used distinctively in AE; but the fact that voicing continues during Zulu /ɓ/ release makes the implosive gesturally quite similar to voiced AE /b/ in both location of supralarygneal constriction and basic glottal setting.

Based on articulatory similarities and differences between the Zulu consonants and the most closely corresponding AE consonants, the following assimilation predictions were made: The lateral fricatives were expected to show two category (TC) assimilation by most AE listeners, as some phonological contrast in English, such as a voiceless apical fricative (e.g., /θ s ∫/, perhaps clustered with /l/) versus /l/ (voiced lateral approximant) or some voiced apical fricative (e.g., /ð z ʒ/, perhaps clustered with /l/), which involve the same articulators (tongue tip and dorsum, glottis), constriction locations (dental/alveolar and posterior constrictions), and constriction degree (fricative) as these Zulu consonants. The velar stops were expected to show a notable category goodness difference (CG) in assimilation to good versus poor AE /k/, that is, to a native consonant involving the same articulatory organs (tongue dorsum, glottis), constriction location (velar), and degree (stop). The bilabial stops were expected to show single category (SC) assimilation as nearly equivalent exemplars of AE /b/ (same organs, constriction location and degree), at least for most listeners. The associated discrimination predictions were that performance would be excellent for the lateral fricatives, quite good but significantly lower for the velar stops, and substantially poorer for the bilabial stops, i.e., TC>CG>SC.

Discrimination of each contrast was tested before assimilation was assessed, to minimize the potential influence that labeling or describing the Zulu consonants may have had on discrimination. Given Werker and Logan's finding (1985) that short-term memory constraints affected English listeners' discrimination of a Hindi contrast that fits the SC definition, but not Hindi listeners' discrimination of the same contrast (native ~ TC), we also assessed whether this factor might influence discrimination more for SC assimilations than for TC assimilations. A difference in memorial influences would further support the differentiation of SC and TC contrasts. It must be noted, however, that neither PAM nor the other two non-native speech perception models (SLM and NLM) make explicit predictions about memorial effects on non-native speech discrimination.

A. Method

1. Participants—The listeners were 22 native speakers of American English (15 female, 7 male) with a mean age of 18.4 years (range=18–20 yr), recruited from an Introductory Psychology subject pool. Participants received course points for their participation. None of the participants had experience with Zulu or any other languages employing the consonant contrasts used in this study. None had a personal or family history of developmental speech, language, or reading disorders. Five other participants were tested but their data were later removed due to ear infection on the test day (*n*=1), delayed language development (*n*=1), or familial speech problems (*n*=1) or reading impairments (*n*=2).[5]

2. Stimulus materials—An adult female native Zulu speaker from Durban, South Africa, was recorded producing multiple tokens of each of the six target consonants in CV nonsense syllable pairs. The syllables used were [ɮɛ]-[ɬɛ] (lateral fricatives), [kʰa]-[kʼa] (velar stops), and [bu]-[ɓu] (bilabial stops).[6] All syllables had high tone on the vowel.[7] The syllables were read aloud individually from a randomly ordered list containing 20 repetitions of each.

The recording was digitized on a VAX 11-780 computer using the Haskins Laboratories' Pulse Code Modulation (PCM) system (Whalen *et al.*, 1990). Individual syllables were extracted and acoustically analyzed by the third author, using a signal analysis program called HADES, which was developed at Haskins Laboratories. The measures included the durations of consonant noise, vowel, and full syllable; rms amplitude of the consonantal noise; VOT (for stops only); spectral centroid values at 15%, 50%, and 85% into the consonant noise; and frequency peaks for *F*0 and each of the first three formants at 15%, 50%, and 85% into the vowel. "Centroid" refers to the spectral center of gravity, or amplitude-weighted mean frequency, calculated as the first moment of a DFT. Centroid values primarily reflect front cavity size and configuration (see Nittrouer *et al.*, 1989). Only those tokens that our Zulu speaker identified in a listening task as unequivocal productions of each category were further considered. Also, any tokens displaying list intonation effects or other odd voice qualities were ruled out as potential stimuli for the perceptual tests. Six tokens were then selected 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 at the 50% point. The first author then independently remeasured the acoustic properties of the selected stimuli (see final values, Table I), using the Signalyze program (Keller, 1994) on a Macintosh computer. Note, however, that centroid values (HADES) could not be computed in Signalyze, and *F*0 and formant measures were added for the first pitch pulse of each stimulus. Note also that final formant values (Signalyze) were based on fast Fourier transforms (FFTs) rather than linear predictive coding (LPC) estimation (as in HADES), and that final *F*0 values (Signalyze) were calculated as the inverse of the period of the glottal pulse nearest to the designated time slice.

In summarizing the acoustic analyses, we use the term "discrete difference" to describe measures that showed no overlap in range of values between the contrasting consonants, and the term "overlapping difference" for measures whose values showed partially overlapping ranges, thus inconsistently differentiating the contrast. For the remaining measures, the range of values for one category was completely subsumed within the range of the other (or nearly

---

[5]Reading deficits were used as an exclusionary criterion because they are often associated with deficient phonological skills (e.g., Scarborough, 1998; Shankweiler *et al.*, 1995). Due to requirements of subject pool use, some screening factors had to be applied *after* subject participation. For the same reason, gender was not balanced in the sample. However, this was not deemed critical, as no sex differences have been reported for speech perception tasks such as those used here.

[6]Different vowels were used for the three contrasts because we also planned to use these stimulus materials for a within-subjects study with AE infants. The vowel difference was deemed necessary to maintain infants' attention across their three required tests (Best *et al.*, 1990).

[7]Zulu is a tone language with a differentiation between high and low tones on syllable nuclei.

so). These latter measures are designated as "no difference" even if the means appear to diverge between the contrasting categories, because of the complete overlap in range.

The acoustic differences for each stimulus contrast are consistent with their productions (e.g., aerodynamic differences). For the lateral fricatives, the discrete differences were that the voiceless syllables showed a higher $F1$ at the first vocalic pitch pulse (M_{diff}=67.5 Hz), a higher $F0$ at 15% into the vowel (M_{diff}=18.9 Hz), and a longer duration of frication (M_{diff}=43.4 ms, or 40% longer than voiced fricatives). The overlapping differences were that voiceless stimuli displayed inconsistently higher mean amplitude (M_{diff}=6.5 rms), higher mean centroid frequency at 50% into the frication (M_{diff}=87.7 Hz), higher $F2$ at 15% into the vowel (M_{diff}=69.5 Hz), higher $F0$ at the first pitch pulse and at 50% into the vowel (M_{diff}=41.4 and 8.5 Hz, respectively), and longer mean duration for the full syllable (M_{diff}=35.3 ms or 11% longer than voiced fricative syllables). The remaining acoustic measures showed no difference. Thus, 3 of 23 measures showed discrete differences; another 6 showed overlapping differences.

For the velar stops, the discrete differences were that ejective release bursts were higher in amplitude (M_{diff}=12.1 rms), longer in duration (M_{diff}=28.8 ms, or 339% longer), and higher in centroid frequencies at all three measured points in the consonant noise (M_{diff}=653.8 Hz) than the bursts of the aspirated stops. These properties are consistent with phonetic descriptions of Zulu ejective velar stops (and our own perceptual observations of them) as somewhat affricated. The overlapping differences were that ejectives also had inconsistently higher mean $F0$ at the first pitch pulse (M_{diff}=28.1 Hz) and lower $F2$ at the first pitch pulse (M_{diff}=90.6 Hz), but higher $F2$ at 15% and 50% into the vowel (M_{diff}=114.3 Hz). All other measures, including VOT, showed no difference. Thus, 5 of 24 measures showed discrete differences, and another 4 showed overlapping differences.

For the bilabial stops, the discrete differences were that the implosives had higher $F0$ and $F1$ frequencies at 15% into the vowel (M_{diff}=54.7 and 23.7 Hz, respectively), and had higher-amplitude bursts (M_{diff}=10.9 rms) and substantial prevoicing as compared to the small, unaspirated voicing lag of the plosives (VOT M_{diff}=72.2 ms). The overlapping differences were that the implosives had inconsistently higher-frequency centroids at 50% and 85% into the consonant noise (M_{diff}=293.1 Hz), slightly shorter vowels (M_{diff}=26.5 ms) but longer full-syllable durations (M_{diff}=32.4 ms), higher $F0$ at 50% into the vowel (M_{diff}=17.5 Hz), lower $F1$ at the first pitch pulse (M_{diff}=13.2 Hz) but higher $F1$ at 50% and 85% into the vowel (M_{diff}=17.1 Hz), and higher $F2$ at 50% (M_{diff}=46.8 Hz) but lower $F2$ at 85% into the vowel (M_{diff}=17.8 Hz). The remaining measures displayed no differences. Thus, 4 of 24 measures showed discrete differences, and another 10 showed overlapping differences.

To summarize, the consonantal portions of all three stimulus sets showed several discrete or overlapping differences between the contrasting phone sets. The velar and bilabial contrasts showed discrete differences in amplitude of consonantal noise; the velar and lateral fricative contrasts showed discrete differences in consonant noise duration which were proportionally much larger for the velars, and the bilabials differed in VOT. The velar contrast also displayed systematic and pervasive centroid frequency differences in the consonant noise bursts, whereas the lateral fricative and bilabial contrasts showed systematic $F0$ and $F1$ frequency differences early in their vocalic sections. The other acoustic measures showed inconsistent or no differences between the contrasting consonants. It is noteworthy that English listeners show low levels of perceptual confusion, across a range of signal-to-noise ratios, for native stop and fricative voicing distinctions, as well as for native affrication differences (i.e., fricative versus stop manner). These observations are most relevant to the Zulu contrasts tested here. By comparison, confusion levels for native place of articulation distinctions are fairly high (Miller and Nicely, 1955). The similarity in number of discrete acoustic differences for our three

nonnative contrasts, together with the classic findings of Miller and Nicely, offer little *a priori* acoustic basis for predicting discrimination differences.

3. Procedure—Listeners first completed a categorial AXB discrimination test for each of the three Zulu contrasts. In this procedure, A and B are tokens of contrasting non-native phonemes. Listeners are told to circle on their answer sheets for each trial whether the middle item (X, or target) is the same syllable as the first or third item. The X is a different physical token than that of the categorially matched A or B item, so that listeners cannot make a simple acoustic identity judgment (e.g., Best *et al.*, 1988; Polka, 1991, 1992). This procedure was used for several reasons: (1) Because the categorial approach asks listeners to determine whether physically different tokens have the same identity or not, it better approximates natural listening conditions than do tasks that present physically identical tokens for judgment (as in "same" trials in AX tasks). (2) Observers display much lower, and easily estimated, response bias in 2AFC (two alternative forced choice) tasks such as AXB than in single-interval decision tasks (e.g., AX), and 2AFC tasks allow measurement of sensitivity to smaller stimulus differences than may be easily assessed with single-interval yes/no tasks such as AX (MacMillan and Creelman, 1991, p. 134). (3) The AXB task was used in the previous investigations of PAM predictions (Best *et al.*, 1988; Best and Strange, 1992).

Each AXB test contained 96 trials in 12-trial blocks (interstimulus interval [ISI] =1 s; intertrial interval=3.5 s; interblock interval=5 s), presented to listeners via audio tape. This was the ISI used in previous PAM reports (Best *et al.*, 1988; Best and Strange, 1992);[8] it should also minimize backward and forward masking between adjacent stimuli. Though this length of ISI might place a load on memory, we assessed this via analysis of short term memory effects (see the next section). The four trial types (AAB, ABB, BBA, BAA) were equally represented for each contrast, and within each test the trial order was randomized. Each of the six tokens per stimulus set occurred four times in each trial type, twice in A and twice in B position, but never paired with the same opposing token more than once.

Following the discrimination tests, a second questionnaire task was conducted involving transliteration of the syllables using English orthography, followed by eliciting additional descriptions, for each set of syllables, in order to evaluate perceptual assimilations. On each trial, the six tokens for a given target syllable were presented. Participants were then directed to write down what the syllable sounded like to them, using English orthography (i.e., "spell as you would in English"), *if and only if* the consonants sounded to them like anything resembling English consonants. The questionnaire then asked them to write any further description they could give regarding the way the stimuli sounded to them [e.g., "it sounded like the speaker was doing —— when she pronounced the consonant" or "it sounded like —— (some nonspeech sound)"]. Participants could not see each others' responses nor discuss their perceptions of the stimuli during the test session.

Listeners were tested in groups of four to six, along one side of a large table in a sound-attenuated room. Stimuli were presented via an Otari MX5050 BQ-II reel-to-reel tape deck connected to a Kenwood amplifier, which fed to a Jamo compact loudspeaker. The speaker was centered on the opposite side of the table, facing the participants (approximately 3 ft from them). Output from the loudspeaker was set to 70±3 dB, as measured from the participants' location.

---

[8]Note that Werker has often used an even longer ISI of 1500 ms in her investigations of non-native speech perception in infants and adults, as has Polka (1991, 1992).

B. Results

1. Discrimination analyses—Percent correct performance was analyzed in a three-way within-subjects analysis of variance (ANOVA) on the factors stimulus contrast (laterals, velars, bilabials)×trial type (whether X matched the first item of the trial [A<u>AB</u> and <u>BB</u>A trials] versus the third [A<u>BB</u>, B<u>AA</u>])×native similarity (whether X matched the more English-like comparison item [A<u>AB</u>, B<u>AA</u>] versus the less English-like one [A<u>BB</u>, <u>BB</u>A]). Trial type provides an index of memory influences. Better performance on trials where X matches the third item of the trial would suggest a recency-type effect, posited to reflect auditory short-term memory constraints (see Crowder, 1971, 1973). Native similarity was determined by considering the phonetic properties of each Zulu phone relative to the closest English phoneme (s), in addition to considering the listeners' assimilations (below). In the case of the Zulu velar stops, the voiceless /k/ is virtually identical to English /k/, whereas ejective /k′/ is obviously more deviant from English /k/. Listeners' assimilations were consistent with this phonetic analysis: they produced more orthographically regular English spellings for Zulu /k/ than for /k′/. The Zulu plosive bilabial has essentially the same pronunciation as English /b/, whereas the Zulu implosive /ɓ/ is less English-like in that it employs a non-English larynx-lowering gesture which results in negative airflow. However, determining native similarity was more difficult for the Zulu lateral fricatives. Both are deviant from English fricatives in terms of place of articulation, both using the same tongue constriction locations (which are similar to AE /l/), and the voicing distinction is virtually identical to that for AE fricatives. However, one apparent basis of difference in English likeness is evident in the listener assimilations. The phonotactically permissible English spellings they wrote for the voiceless lateral fricative (/s ʃ tʃ sl ʃl/) have a higher mean frequency of occurrence in word-initial position ($M_{frequency}$=0.0166), according to the Francis and Kuçera database (1982), than did their permissible spellings (/l z/) for the voiced lateral fricative ($M_{frequency}$=0.0037). Given the lack of other bases for deciding, the voiceless lateral fricative was designated the more English-like item for the native similarity factor.

The main effect of contrast was significant, $F(2,42)$ =178.91, p<0.0001. Tukey tests revealed that discrimination for the lateral fricatives was significantly better (M = 95% correct, s.e. =0.49) than for the velar stops (M = 89.4%, s.e. =1.4), which was significantly better than for the bilabial stops (M=65.9%, s.e. =1.5) (all p's<0.01). Nevertheless, even for the bilabials discrimination was significantly above chance (50% correct), $t(21)$ =11.59, p < 0.001.

The main effect of trial type was only marginally significant (p=0.08). However, the trial type×contrast interaction was significant, $F(2,42)$ =5.495, p<0.008. Simple effects tests revealed that trial type was significant only for the bilabial test, $F(2,21)$ =7.29, p<0.01, with performance higher on recency-type trials (M=69.9%, s.e. =1.9) than on primacy-type trials (M=61.9%, s.e. =2.1). This suggests that auditory memory influenced discrimination of the bilabials, but not of the velars nor of the lateral fricatives, which failed to show recency effects. Nonetheless, discrimination of the bilabials was significantly above chance for both recency-type trials, $t(21)$ =10.57, p<0.0001, and primacy-type trials, $t(21)$ =5.57, p<0.001 (see Fig. 1).

The native similarity main effect was also significant, $F(1,27)$ =34.41, p<0.0001. Discrimination was significantly better when the target (X) was more English-like (M=86.79%, s.e.=1.78) than when it was less English-like (M=80.13%, s.e.=1.81) (see Fig. 1). Although the contrast ×native similarity interaction was nonsignificant, we ran a simple effects test on it in order to determine whether the similarity effect was significant for each contrast individually. The effect was indeed significant for each contrast: SC, $F(1,42)$=7.296, p<0.013; CG, $F(1,42)$ =47.518, p <0.0001; TC, $F(1,42)$=29.2, p<0.0001. To determine whether the effect differed in magnitude among the three contrasts, we then calculated difference scores (more-English-like minus less-English-like) and conducted a contrast×trial type ANOVA. No main effects or

interactions were significant in this analysis, and Tukey tests among the contrasts were all *ns*, indicating a lack of variation in magnitude of the native similarity effect.

Although response bias is low for AXB and other 2AFC discrimination procedures, we applied MacMillan and Creelman's recommended bias-correction procedure to the percent correct data (1991, p. 127). The formula, $q_{2AFC}=[p(c)_{2AFC}-0.5]/(1-0.5)$, yields the proportion of guessing-corrected performance above chance, which we multiplied by 100 to obtain corrected percent above-chance performance. Since this is a linear transformation of the raw percentage data, the ANOVA results were identical to those for the uncorrected scores. Both the uncorrected and corrected cell means are listed in Table II.

2. Assimilation patterns—English spellings and descriptions of the Zulu consonants by each participant, on each contrast, were categorized according to whether the participant used the same or different consonant spellings for the contrasting syllable onsets, as well as whether their additional written descriptions identified any differences they noticed in the productions or sound of the consonants. If both consonant onsets were spelled identically, or were phonologically equivalent in English orthography (e.g., CA and KA), and the participant's additional descriptions failed to note any other consonantal differences,[9] the participant's assimilation pattern for that contrast was categorized as SC. If instead the contrasting consonants were spelled with a common letter, yet one member of the pair was further modified by punctuation marks or by additional letters to emphasize some phonetic feature (e.g., K followed by H to indicate aspiration), and/or the participant's written description noted some phonetic (or acoustic) discrepancy between the two consonants, then the assimilation of that contrast was considered a CG difference within a single English consonant. That is, a perceived goodness difference was inferred from the discrepant notation/ description. But if the two consonant onsets were spelled with different letters or combinations of letters that indicate phonologically different English pronunciations, the assimilation pattern was categorized as a TC type. If the spelling and description had referred to a stimulus set as falling somewhere in between two or more English consonants (e.g., "between 'sh' and 'th'" or "sometimes sounds like 's' sometimes like 'sh' or 'zh'") for one or both Zulu consonants, then the pattern would have been categorized as UC (uncategorized-categorized) or UU assimilation, respectively. Alternatively, if the listener gave a name or description only of some nonspeech sound (e.g., "snapping" or "popping sound" or "whooshing"), it would have been designated as a NA type. No participants indicated that any of the Zulu consonants were heard as uncategorized speech sounds or as NA nonspeech sounds. That is, all were described as English consonants or consonant sequences.

All 22 participants showed the expected TC assimilation of the lateral fricatives to some phonological distinction in English. Each labeled the voiceless lateral fricative as some AE voiceless fricative or affricate involving the same articulators (tongue tip/body); ten combined this with /l/, /h/, /t/ or /z/. For the voiced lateral fricative, ten gave the label "l," five gave "z," three wrote a voiced fricative combined with other fricatives involving tongue tip/body, and the remaining four provided clusters of voiced fricative + "l" (thus combining the same articulators, constriction locations, degree, and/or laryngeal setting) (see Table III). A few participants provided additional articulatory descriptions, most involving tongue tip/body constrictions (silent "n," unpronounced "l," soft "c;" stronger "s" or "l"). Only one participant offered a more acoustic-oriented description, indicating a "slight click on the 'l'" he had heard for the voiced fricative.

---

[9]Comments that instead identified vowel qualities or intonational properties do not reflect perception of the consonants *per se*, and so were not factored into the consonant assimilation determinations.

For the velar stops, all participants were again consistent in their assimilations, this time reporting the expected CG difference in goodness of fit to AE /k/. All listed /k/ as their primary response (same supralaryngeal articulator, constriction location, and degree, and same laryngeal gesture), but all noted the ejective with a q, c, g, ch, or h or a mark such as an apostrophe or a dash following the /k/. All but 4 participants also wrote further descriptions of the ejective, with 16 indicating that the ejective included some unusual articulation in the throat (pharynx) (choke; gagging; gurgle; throaty; clearing throat; in back of throat), and/or involving the tongue tip/body (clucking; clicking [painful; nasalized; at roof of mouth; Bushman-like]), and two giving more acoustic-oriented descriptions (slight clacking noise; broken up).

The bilabial stops yielded a somewhat less consistent assimilation pattern, although about $\frac{2}{3}$ of the participants (n =15) showed the expected SC assimilation to a single AE consonant lacking any notated differences in goodness of fit. These SC listeners reported both bilabial consonants as /b/ (same articulatory organ, supralaryngeal constriction location, and constriction degree) without additional spelling, marking, or descriptive differences; however, a subset of these did report vowel or intonation differences (n=9). Discrimination performance for SC listeners overall was poor (M=64.9%, s.e.=1.95); it was no higher for the subset who noted vowel or intonation differences (M=63.93%, s.e. = 2.55) than for those who did not (n=6: M=65.87%, s.e. =2.52). Two other participants showed a CG assimilation pattern to /b/, providing added articulatory descriptions for the implosive (harder; "mb" described as softer /b/). Their discrimination performance (M=72.23%, s.e.=2.33) was better than that of the SC participants. Four others showed TC assimilation as /b/ vs /v/ (different constriction location and degree); however, their discrimination (M=65.91%, s.e.=2.25) was no better than that of the SC participants. The remaining participant failed to describe one Zulu bilabial; his assimilation was not classifiable.

3. Discrimination reevaluated—Given the individual variations in assimilation of the bilabial contrast, we tested whether the discrimination results would be upheld for just the 15 participants who had shown the predicted assimilation types on all three contrasts: TC assimilation of the lateral fricatives, CG assimilation of the velar stops, and SC assimilation of the bilabial stops. The results remained essentially the same as for the full group. The main effect of contrast was significant, $F(2,28)$ =114.01, p<0.0001, supporting the predicted performance pattern of TC>CG>SC. The trial type main effect remained nonsignificant, while the contrast×trial type interaction became marginal, $F(2,28)$=2.805, p<0.08. However, simple effects tests again indicated an advantage on recency-type trials for the bilabial contrast, $F(1,14)$=4.41, p=0.05, but not for the fricatives or velars. Bilabial discrimination remained significantly above chance both for recency-type trials (M=68.67%, s.e.=2.36), $t(14)$=7.92, p<0.0001, and primacy-type trials (M=60.74%, s.e.=2.83), $t(14)$=3.8, p <0.002. The main effect of native similarity also remained significant, $F(1,14)$=18.99, p<0.0007, and did not interact significantly with trial type or contrast. Thus, listeners who showed the predicted assimilation for all three contrasts performed better on all of them when X was more English-like.

C. Discussion

The assimilation results are largely consistent with the PAM predictions made on the basis of articulatory-phonetic similarities between Zulu and AE consonants. As expected, the lateral fricatives were assimilated as a TC contrast, and the velar stops as a CG difference within a single English consonant, by all listeners. Over $\frac{2}{3}$ of the participants also showed the predicted SC assimilation of the bilabial stops, reporting no differences in the consonants' goodness of fit to AE /b/. The remainder, who showed either CG assimilation of the bilabials to /b/ or TC assimilation to /b/ vs /v/ or /w/, displayed clear responsiveness to articulatory properties, in that they always reported hearing consonantal constrictions involving lips as the articulator, distinguished either by a noncontrastive difference in degree of constriction (e.g., "more pursed

or tense") or by a phonotactically permissible AE contrast in constriction degree and/or location (/v/). Most listeners' assimilations referred to articulatory properties of the stimuli; that is, the listeners seem to have approached the task as "naive phoneticians," with a focus on articulators, constriction location, and degree.

The variations in discrimination across the contrasts, considered in light of the assimilation patterns, also supported the PAM prediction of the performance pattern TC >CG>SC. That is, the lateral fricatives were discriminated better than velar stops, which were discriminated better than bilabial stops. By comparison, the Zulu click contrasts tested by Best and colleagues (1988), which had yielded clear NA assimilation, were discriminated between 80.6% and 99.1% correct. Thus, click discrimination ranged between the TC and CG levels we found here, and was substantially better than the SC level, again in keeping with PAM predictions.

The native similarity effects with AE listeners are of particular theoretical interest. That discrimination performance was influenced by the native likeness of the target item in the AXB trials is consistent with the PAM claim that perceivers are sensitive to variations of a native consonant. For the CG assimilation case, this effect may appear to be consistent with NLM predictions (Grieser and Kuhl, 1989; Kuhl, 1992; Kuhl *et al.*, 1992) about goodness-related discrimination effects (cf. Miller, 1994; Volaitis and Miller, 1992). Specifically, NLM predicts that discrimination should be worst among tokens that are acoustically similar to the prototype, and best among tokens that are non-prototypical, of the same native category. However, our finding seems to show the opposite pattern—*better* discrimination performance for more nativelike targets (i.e., more prototypical) and *poorer* performance for less nativelike (non-prototypical) ones. Possibly, methodological differences contribute to this apparent reversal of NLM findings. We employed a categorial AXB task whereas Kuhl and colleagues tested detection of stimulus changes against a repeating background. Perhaps the perceived equivalence between target and matching items in our categorial AXB task are greater when the target is more English-like (i.e., corresponds to an NLM prototype) than when it is less English-like (i.e., corresponds to a non-prototype). This might provide an NLM-compatible interpretation of native similarity effects (see also Polka and Werker, 1994). However, the native similarity effects for SC and TC assimilations are inconsistent with NLM expectations, given that these types involve a notable difference in goodness of fit to the associated native phoneme. That is, NLM should expect a native similarity effect in discrimination only for the CG contrast, with significant differences in magnitude of the effect between the CG contrast and the other two types. This expectation was not supported; all contrasts showed the effect, and its magnitude did not differ significantly among them.

Two aspects of the Zulu bilabial findings must also be addressed: above-chance discrimination by listeners who showed SC assimilation, and poor discrimination by those who showed TC assimilation. SC listeners' discrimination was poor, as predicted, but was nevertheless significantly above chance, even according to the bias-corrected scores. This may perhaps seem unsurprising, in light of ample evidence from studies of categorical perception that within-category discrimination is usually significantly better than chance. But the important question is, why isn't it at chance, specifically in the present SC case? These listeners had failed to detect any sort of phonological contrast, or even any differences in phonetic goodness of fit to AE / b/. Obviously, whatever remaining properties they detected did not support very good discrimination. But what actual stimulus differences might they have heard? There was a reliable difference in voicing between the unaspirated /bu/ and prevoiced /ɓu/; however, both voicing values are found in allophones of AE /b/ and are difficult for English listeners to discriminate (Lisker and Abramson, 1967), and few participants reported such differences. On the other hand, nine listeners reported differences in vowel quality or intonation, perhaps associated with slight differences in $F1$ and $F2$ onsets, and mid-vowel $F0$ differences,

respectively, for /bu/ vs /ɓu/. Still, those who reported such differences discriminated the Zulu bilabials no better than those who did not.

The recency-type effect for the bilabial contrast alone may suggest another clue, although as noted earlier, neither PAM nor NLM and SLM make *a priori* predictions about auditory memory effects on discrimination. Discrimination in cases where the listener fails to detect either a phonological contrast or a phonetic goodness difference would be expected to involve detection of nonlinguistic auditory differences, and thus to show an influence of auditory memory reflected in a recency effect (see Crowder, 1971, 1973). Recency effects in discrimination would not be expected for CG or TC assimilation, which involve detecting phonetic or phonological differences, respectively, rather than nonlinguistic differences. That is, we speculate that detection of contrastive phonological distinctions versus non-contrastive phonetic details versus nonlinguistic auditory properties is somehow differentiated in non-native speech perception. Although this three-way division is superficially consistent with Werker and Logan (1985)'s proposal for separate phonological, phonetic, and auditory processing levels, our own view more closely follows the direct realist position that listeners detect information in signals about the nature of the event that produced the signal. In the case of speech signals, listeners could detect several types of event information: articulatory patterns that signal phonological distinctions in a language, articulatory patterns that are noncontrastive phonetic variants of phonemes in the language, or nonlinguistic aspects of vocal (or other) sound-producing events such as breathiness, emotional intonation, murmuring, clacking noise, choppiness, etc.

Consistent with the preceding reasoning, recency effects were found for SC but not for CG or TC assimilations (see Fig. 1). However, the difference in mean discrimination levels for the three contrasts raises the possibility that the recency effect for bilabials is due simply to the generally poor performance level rather than to the detection of nonlinguistic, as opposed to phonetic or phonological, information *per se*. However, we can assess this possibility by testing for recency effects in discrimination of the nine NA (nonassimilable) Zulu click contrasts examined by Best and colleagues (1988). Performance on those clicks was higher than the current SC discrimination, and comparable to CG and TC discrimination levels in the present study (80%–99% correct). Yet discrimination of the clicks apparently involved detection of nonlinguistic rather than phonetic and phonological differences, as with the SC contrast in the current study. Therefore, we reexamined the click discrimination data in a new ANOVA on trial type×feature type (voicing contrasts versus place of articulation contrasts)×phonetic contrast (voicing contrasts: prevoiced/short-lag unaspirated, unaspirated/long-lag aspirated, prevoiced/aspirated; place contrasts: dental-lateral, lateral/palatal, dental/palatal). Only trial type was significant, $F_{(1,8)}$=22.71, p<0.002, indicating a recency effect: better discrimination for trials in which the target matched the third item of the trial (M=92.95%, s.e. =0.77) rather than first item (M=89.92%, s.e.=1.03). So, recency effects appear to be associated specifically with detection of nonlinguistic as opposed to phonological or phonetic differences, rather than being associated with poor discrimination.

There is another puzzle, however. The small number of listeners who reported TC assimilation of the bilabials showed *poor* discrimination, no better than the SC listeners. Why? We suspect the reason that these unexpected and infrequent cases of TC assimilation for the bilabials, unlike the expected and unanimous TC cases for the lateral fricatives, showed an assimilation-discrimination discrepancy may be attributable to task order. We had listeners complete the discrimination task prior to the spelling/description task in order to minimize influences of categorization on discrimination performance; this was a necessary experimental control for evaluating PAM hypotheses about the influence of perceptual assimilation patterns on discrimination. However, this minority of listeners may have felt compelled to generate some AE phonological distinction when presented with the two bilabial categories in the second task,

even though their poor AXB performance strongly suggests that they had not detected any such differences during the preceding discrimination task.

The more pervasive TC effect, however, was the predicted one of excellent discrimination in the case of unanimous TC assimilations of the lateral fricatives. Here, the order of experimental tasks was crucial for ruling out any possibility that categorization experience with the stimuli could have directly affected discrimination performance within the experimental context. This TC assimilation pattern with near-ceiling discrimination is the most surprising of the PAM predictions, from the perspective of classic reports that adults have serious difficulties in labeling and discriminating nonnative phonetic contrasts. Moreover, this TC pattern has received the least prior research attention, having been reported only by Best and Strange (1992), and there only for categorical perception of a synthetic continuum rather than of multiple natural utterances. For these reasons, we conducted a second experiment to extend our investigation of TC assimilation to another non-native contrast from a different language.

III. EXPERIMENT 2

For this study, we chose a stop consonant contrast from a second African language, Ethiopian Tigrinya, which is from a different language family (Afro-Asiatic: Semitic: Ethiopic) than Zulu (Niger-Kordofanian: Niger-Congo: Bantu) (Ruhlen, 1975). The contrast was between the ejective bilabial versus alveolar stops /p′/ and /t′/. The consonants contrasted in constriction locations that occur in AE rather than in laryngeal gestures, as in experiment 1; the laryngeal gesture of both was non-native to English. We also tested discrimination of two native AE fricative voicing contrasts involving phonemes that had appeared in the experiment 1 participants' spellings of the Zulu lateral fricatives: /s/-/z/ and /ʃ/-/ʒ/ and involved tongue tip/ body as the active articulators, for comparison to the results with that contrast. Because the lateral fricatives had been presented with the lax vowel /ɛ/ in open CV syllables, which is phonotactically impermissible in English, we used the same vowel and CV context in all contrasts tested in experiment 2. To directly compare the results to those for the lateral fricatives (TC) of experiment 1, we used the same testing procedures.

A. Method

1. Participants—The listeners were 19 native speakers of American English (10 female, 9 male) with a mean age of 18.7 years (range=18–20 yr). None had experience with Tigrinya or any other languages employing ejective consonants. None had a personal or family history of developmental speech, language, or reading disorders. Eight other participants were tested but their data were removed from the final data set due to developmental and/or familial speech impairments (*n* =5), familial language disorders (*n*=1), or reading impairments (*n*=1) (see footnote 5), or chance-level discrimination of the English control contrasts (*n*=1).

2. Stimulus materials—A male native Tigrinya speaker from Ethiopia (Eritrea) was recorded producing multiple tokens of each of the two non-native CV nonsense syllables /p′ ɛ/-/t′ ɛ/. The supralarngeal articulators, constriction locations, and constriction degrees for these two stops correspond to those of the AE voiceless stops /p/-/t/; however, the ejective laryngeal gesture of both is not used in English. The syllables were read aloud individually from a randomly ordered list containing 20 repetitions of each. The AE contrasts /sɛ/-/zɛ/ and / ʃɛ/-/ʒɛ/ were recorded according to the same procedure by a female native AE speaker (author CTB).[10] The recordings were digitized and analyzed as in experiment 1. Six tokens were selected per category, matched as closely as possible between the contrasting syllables of each

[10]We also recorded, and collected perceptual data, for the same AE C's followed by /eⁱ/; data on the latter stimuli will not be reported here because the vowel environment differed from the Zulu lateral fricatives and the Tingrinya stimuli. However, the perceptual results for those stimuli were virtually identical to those for the AE stimuli reported in experiment 2.

pair for overall duration, fundamental frequency and contour, and vowel formant frequencies (see Table IV).

The discrete differences between the final stimulus sets for the contrasting Tigrinya ejectives are in the spectrum, duration, and amplitude of the release bursts. The /t'/ bursts were longer in duration (M_{diff}=8.6 ms, or 72% longer), higher in amplitude (M_{diff}=9.9 rms), higher in centroid frequencies throughout (M_{diff}=986.3 Hz), and had higher $F2$ and $F3$ values at the first pitch pulse of the vowel (M_{diff} =516.0 and 1182.2 Hz, respectively) than the /p'/ bursts. All other acoustic measures on the syllables, including VOT, showed no difference between the two categories. Acoustic measures for the AE fricative contrasts were quite similar between paired stimulus sets, except for the obvious voicing difference.

3. Procedure—As in Experiment 1, listeners first completed categorial AXB discrimination tests for the Tigrinya and AE contrasts. Following the discrimination tasks, listeners completed the assimilation questionnaire for each set of Tigrinya syllables, as in experiment 1. Listeners were tested in groups of four to six in the same experimental setup as before.

B. Results

1. Discrimination analyses—The AXB discrimination data were submitted to ANOVA for the within-subject effects of contrast (the two AE contrasts versus the Tigrinya contrast) ×trial type (primacy versus recency).[11] This time, the only significant effect was contrast, $F_{(2,36)}$=6.792, p<0.003. Discrimination performance was essentially at ceiling for the two AE contrasts (for /s/-/z/, M=98.8% correct, s.e.=0.33; for /ʃ/-/ʒ/, M =98.8% correct, s.e.=0.37), but was somewhat lower and more variable, though still excellent, for Tigrinya /p'/-/t'/ (M=91.4% correct, s.e.=2.02). The trial type effect and the interaction were nonsignificant. Cell means for percent correct, as well as for bias-corrected percent performance above chance, are shown in Table V.

Performance on the lateral fricatives of experiment 1 was compared to performance for each of the AE fricative voicing contrasts in between-subject contrast×trial types ANOVAs. Discrimination was significantly lower, though still excellent, for Zulu /ʒe/-/ɬe/ (M=95% correct, s.e. = 0.49) as compared to both AE /s/-/z/, $F_{(1,39)}$=26.243, p <0.0001, and AE /ʃ/-/ʒ/, $F_{(1,39)}$=32.14, p<0.0001 (see preceding paragraph for AE means).

We also directly compared performance on the two non-native TC contrasts in a between-subject contrast×trial type ANOVA. There were no significant differences.

2. Assimilation patterns—Assimilation patterns were determined according to the experiment 1 criteria. As predicted, the great majority of participants assimilated the Tigrinya ejectives as a TC contrast (n=16). Of the 16 who showed TC assimilation, 12 reported hearing /p/-/t/, that is, their assimilations were consistent with the supralaryngeal articulators, constriction locations, and constriction degree. Consistent with the notion that listeners can detect within-category phonetic differences, i.e., between non-native phones and the native categories to which they are assimilated, some listeners noted deviant articulatory details involving throat and/or larynx for the /p/-/t/ assimilations (e.g., click in the throat; windy—a lot of breath behind it; swallowing the consonant; abruptly cut off; sucked in); others noted deviant supralryngeal articulations (spitting out the syllables; hard or pronounced P and T). Two other TC listeners reported /p/ vs /pt/ or /pb/, and two reported an isolated vowel (i.e., no consonant) vs vowel_t. Two of the remaining showed SC assimilation, one reporting

---

[11]Native similarity was not included as a factor because neither the non-native contrast nor the native contrasts of experiment 2 involved differences in English-likeness.

"EH"-"EH," the other /p/-/p/. Consistent with PAM expectations, the latter 2 participants showed substantially lower discrimination (*n*=2, *M*=64.58%, s.e.=13.22) than the 16 who displayed TC assimilation (*M*=94.09%, s.e.=1.96). The final participant failed to describe one Tigrinya consonant; his assimilation was not classifiable.

C. Discussion

The findings from experiment 2 are straightforward. The TC assimilation pattern and its associated high level of discrimination clearly generalized to another non-native contrast. Experiment 2 involved a different type of non-native phonetic contrast, and a second unrelated language, than in experiment 1. Thus, the TC assimilation pattern apparently applies to constriction location contrasts as well as laryngeal gesture contrasts (voicing).

Note, however, that the TC assimilation of non-native contrasts reported in both experiments yielded modestly but significantly lower discrimination (low-mid 90% range) than do comparable native contrasts (near-100% range). Listeners appear to be sensitive, simultaneously, both to information that may be relevant to a native phonological contrast, and also to articulatory differences between nonnative phones and the most similar native phonemes.

Finally, as with the Zulu bilabial results in experiment 1, some individual differences were apparent in assimilation and discrimination of at least some non-native consonant contrasts. Two experiment 2 participants failed to note any phonetic or phonological differences between the Tigrinya ejectives, showing SC assimilation, with concomitantly poor discrimination.

IV. CONCLUSIONS

The results of the reported experiments support the notion that listeners perceptually assimilate and discriminate non-native consonants with respect to their phonetic similarity to native contrasts, in accordance with predictions from the perceptual assimilation model (PAM: Best, 1994a,b, 1995; Best *et al.*, 1988). Specifically, for non-native contrasts in which listeners perceived a correspondence to some native phonological distinction, deemed as two-category (TC) assimilation, discrimination was excellent—above 90% correct. This TC pattern was evident for both a laryngeal gesture distinction and a constriction location distinction from two unrelated languages. By comparison, when the contrasting non-native consonants were heard as differing in goodness of fit to a single native consonant, indicating CG assimilation, discrimination was very good though significantly lower than in TC cases. Thus, while listeners detected variations in the details of items they perceived as variants of a single native consonant, this did not benefit discrimination as much as did the detection of phonologically contrastive information. Finally, when listeners perceived a non-native contrast as equally good variants of a single native consonant, displaying SC assimilation, discrimination was much poorer, as expected. The full set of findings is highly supportive of PAM's proposal of systematic relations between assimilation and discrimination, confirming the discrimination order of TC>CG>SC.

The assimilations of non-native phones to AE consonants corresponded well, for nearly all listeners, to the predictions we had developed from principles of articulatory phonology (see experiment 1 introduction; cf. Browman and Goldstein, 1986, 1989, 1990a,1990b, 1992). Those predictions focused on the use of the same articulators, constriction locations, and/or constriction degrees by non-native and native consonants. Consistent with those expectations, listeners assimilated the Zulu voiced bilabial stops to AE voiced consonants involving the same articulator, that is, lip gestures. Typically, the same location and constriction degree were involved (bilabial stop /b/), though sometimes constriction location and degree differed (labio-dental fricative /v/). Similarly, listeners assimilated the Zulu voiceless and ejective velar stops to the AE voiceless velar stop (/k/), thus to the same articulators (tongue dorsum and glottis)

and the same supralaryngeal constriction location and degree, although glottal constriction degree differed for the ejective /kʹ/. The voiced and voiceless lateral fricatives strongly tended to be assimilated to the AE lateral approximant /l/ (same articulators and constriction locations, but different constriction degree), often combined with voiced or voiceless apical fricatives, respectively (same articulators and constriction degree). And the Tigrinya bilabial versus alveolar ejectives tended to be assimilated to AE voiceless stops using the same supralaryngeal articulators, constriction locations and degree (/p/-/t/).

At the same time, participants often noted additional phonologically irrelevant articulatory features in their written descriptions of the stimulus sets, distinguishing the non-native consonants from the native consonants they perceived as most similar. To illustrate, in addition to their English spellings, listeners sometimes described vocal tract sounds resulting from constrictions of the involved articulators and locations, including tongue body and pharynx (choking, gagging, gurgling, throat-clearing, throaty, guttural sounds), tongue tip (clucking, clicking, stronger /s/), lips (lip-pursing, [lip] muscle-tensing, harder /b/), etc. They only rarely noted nonarticulatory nonspeech sound properties such as clacking noise, broken up.

Three other findings offer several additional insights about nonnative speech perception. First, the native similarity effect, an asymmetry favoring discrimination when the target (X) is the more rather than less native-like member of a nonnative contrast, was found for TC, CG, and SC assimilation types alike in experiment 1. This suggests that familiarity with the typical phonetic form of native consonants aids rather than hinders discrimination, whether the listener is attending for information about phonological contrast, or phonetic goodness of fit to a single phoneme, or nonlinguistic stimulus variations. This suggests that native speech experience results in more stable perception of tokens that are more nativelike, regardless of overall performance level or type of information being discriminated. Evidence of the converse, that perception of less nativelike utterances is less stable, can be seen in the common experience that perception of foreign-accented utterances in the listener's L1, or of utterances in a late-learned L2, is more effortful and error-prone than perception of native L1 utterances.

A second, and perhaps related, finding is that although discrimination of TC contrasts was quite high and that of the SC contrast was quite low, performance differed significantly from ceiling and from chance, respectively. Both observations suggest that listeners retain greater sensitivity to articulatory-phonetic variants of non-native consonants, i.e., show lower perceptual stability or lower perceptual equivalence among tokens, than they do for for native consonants. This burdens discrimination somewhat in TC cases, where the parallel to native phonological contrasts should otherwise have yielded ceiling performance, but it *aids* discrimination in SC cases, where *lack* of correspondence to a native phonological contrast or to a phonetic difference in goodness of fit should otherwise have yielded chance performance. These discrepancies from the upper and lower performance extremes indicate that listeners detect not only the presence/ absence of phonological contrast, but also detect phonologically irrelevant phonetic and/or nonlinguistic details. This finding, together with the native similarity effect, appears compatible with the notion discussed in experiment 1, that listeners are able to discriminate three types of information in speech: phonological, phonetic, and nonlinguistic (see also Hallé *et al.*, 1999; Hallé *et al.*, 1998, 2000; Whalen, 1984, 1991). As argued earlier, this could simply involve detection of certain types of information, and need not entail three qualitatively different cognitive processes (see Werker and Logan, 1985).

Third, only the SC contrast elicited recency effects in discrimination. This finding suggests a qualitative division between detection of linguistic (phonological, phonetic) and nonlinguistic information in speech. This memory effect, putatively auditory, occurred only when listeners failed to report hearing phonological or phonetic differences, presumably leaving only a nonlinguistic basis on which they could have discriminated. The implied relationship between

recency effects and nonlinguistic auditory discrimination was supported by a reanalysis of earlier findings with Zulu clicks, which had been perceived as nonassimilable (NA) nonspeech sounds (Best *et al.*, 1988). Like the SC bilabials, the clicks were also discriminated on a nonlinguistic rather than a phonological or phonetic basis; however, they were discriminated much better than the bilabials. The new analyses revealed a recency effect for click discrimination. The apparent restriction of recency effects to SC and NA cases further supports the differentiation of discrimination of nonlinguistic versus phonological and/or phonetic information in speech. If no such difference existed, the recency effect should have been found across all assimilation types.

While we have interpreted our findings in terms of the PAM model, the differences in TC, CG, and SC discrimination may be, in some ways, reminiscent of classical findings on categorical perception (CP) with synthetic speech continua. Although the original CP claim had been that listeners discriminate speech stimuli only so well as they identify or label them differently, much evidence has indicated that within-category discrimination is usually significantly better than that predicted by labeling functions. In particular, discrimination of tokens near the category boundary, i.e., inconsistently labeled or ambiguous tokens, is above chance (cf. Best *et al.*, 1981). Thus, listeners typically display some sensitivity to within-category variations, though certainly less than that for between-category differences. Those observations might be extrapolated to the better discrimination of TC than CG contrasts, and of CG than SC contrasts. It is important to note, however, that CP findings typically involve unnatural synthetic stimulus variations, whereas the present research and other non-native speech studies involve perception of multiple tokens of natural utterances, a situation that better approaches the natural conditions involved in listening to and learning unfamiliar languages.

It is also important to consider whether and how the findings relate to the other non-native speech models discussed earlier: the speech learning model (SLM): Flege, 1986, 1989, 1995) and the native language magnet model (NLM: Grieser and Kuhl, 1989; Kuhl, 1991, 1992; Kuhl *et al.*, 1992). The SC findings might be seen as consistent with the NLM claim, and the SLM implication, that two non-native phones which are quite similar to a native phoneme should each be difficult to discriminate from it. By extension, they should also, presumably, be difficult to distinguish from each other. Both models might also be extended to account for the TC findings, as a case of non-native phones that are easy to discriminate because they are similar to two different native phonemes, and thus virtually identical to a native contrast. That account, however, is indistinguishable from PAM's explicit hypotheses about TC assimilations. Turning to CG assimilation, the results appear consistent with NLM (and SLM) in showing systematic differentiation of good (closer) versus poor (more distant) examples of a given native phoneme. The native similarity effect of experiment 1 is particularly relevant to NLM claims about asymmetries in discrimination of good versus poor exemplars of a native phoneme. However, as discussed there, it is uncertain whether the direction of our native similarity effect supports or conflicts with NLM. Further research would be needed to determine this. Moreover, NLM should predict a discrimination asymmetry for CG but not for TC or SC assimilations, yet significant asymmetries were observed for all three types of contrast. Thus, native speech experience aids categorial discrimination not only when a non-native contrast assimilates to a phonetic goodness difference within a native phoneme (CG), as emphasized in NLM predictions, but also when listeners fail to detect goodness differences and hear only some nonlinguistic difference (SC), as well as when they detect some phonological distinction (TC).

While certain results can be interpreted *a posteriori* as being compatible with both models, a more fundamental caveat with respect to how well NLM and SLM can address the present findings is that both models focus on the attributes of individual phonetic categories. PAM instead focuses on the functional organization of the native phonological system, specifically on the phonological distinctions between, and phonetic variations within, native phonological

equivalence classes. Importantly, neither SLM nor NLM would have generated the current set of comparisons. And neither offers a singular, coherent account of the findings like PAM does. Thus, a key contribution of PAM is its provision of a theoretical motivation for systematic comparisons among diverse types of non-native contrasts within the broader context of phonological systems.

Another important theoretical issue, not directly examined here, is how native language effects on non-native speech perception emerge developmentally. Infant research indicates that some native language influences appear during the second half-year, with declining discrimination of at least some non-native consonants by 8–10 months (e.g., Best *et al.*, 1995; Werker, 1989; Werker *et al.*, 1981; Werker and Lalonde, 1989), and of some non-native vowels by 6–8 months (Polka and Werker, 1994; cf. Kuhl *et al.*, 1992; but see Polka and Bohn, 1996). Interestingly, there is no developmental decrease for nonnative Zulu click consonants, consistent with AE adults' very good discrimination and assimilation of them as NA nonspeech sounds (Best *et al.*, 1988, 1995). Even more intriguing, however, is that older infants' perception of both native and non-native speech still differs from that of adults in several nontrivial ways, suggesting that they do not yet perceive phonological contrasts like adults (e.g., Best 1991; Hallé and de Boysson-Bardies, 1996; Stager and Werker, 1997). Those developmental differences have led some to posit that infants are initially responsive to language-universal properties of speech, then begin to recognize language-specific phonetic patterns, and only later discover the contrastive phonological functions of native phonetic classes, perhaps in relation to increases in size of their early lexicon (e.g., Best, 1993; Stager and Werker, 1997; Werker and Pegg, 1992). Against the backdrop of our discussion of the three types of information that adults detect in speech (phonological, phonetic, nonlinguistic) we suggest that infants progress developmentally from detection of only nonlinguistic (or perhaps nonspecific phonetic) information in speech, to recognition of how phonetic variants fit into (or fail to) language-specific phonetic classes, to eventually discovering the phonologically contrastive functions those phonetic classes serve in distinguishing native words. Further research on infants' changing perceptions of diverse nonnative contrasts, assimilated by adults as TC vs CG vs SC (vs NA) contrasts, will be needed to test those speculations. We note, however, that the proposed developmental path is consistent with the classic direct realist view of Gibson and Gibson (1955) that perceptual learning involves the increasing differentiation of the lawful stimulus information provided by real-world events. In the case of speech, this differentiation is posited here to involve the emerging recognition of classes of articulatory gestures employed in native speech, followed by discoveries about how those gestural classes help to distinguish among native words.

Before closing, we must address some limitations of the present investigation. The primary methodological limitation lies in our assessment of assimilations. Having listeners give a single native spelling and description of each stimulus set, following the discrimination task, may bias them to search for some between-set difference they were not attending to in the discrimination task. For a more refined approach, forced-choice or perhaps even open-set spellings could be obtained for each token in an assimilation task involving multiple, randomized repetitions. Such data could be easily subjected to standard statistical analyses. Additionally, listeners could rate the goodness of fit between each nonnative token and their associated native-language spelling. Such ratings would be especially useful for differentiating between CG and SC assimilation patterns (see Best *et al.*, 1996; Calderón and Best, 1996). However, other task adjustments would be needed to evaluate listeners' perception of nonlinguistic properties.

Several other aspects of assimilation also deserve further examination, including the basis for predicting the most likely assimilations of a given nonnative contrast by listeners of a given language community. Another finding that calls for further study is the striking individual

variation in assimilation patterns for some non-native contrasts, as we found for Zulu bilabials. Evidently, the phonetic properties of non-native phones reflect multiple dimensions of similarity to various native phonemes, and listeners may differ in their attention to specific dimensions.

Additional research is also needed to substantiate the proposed differences in perception of phonological, phonetic, and nonlinguistic information in non-native speech. For example, native phonotactic rules (phonological), such as constraints on the vowels permitted in open and closed syllables, versus native coarticulatory patterns (phonetic) such as anticipatory or carryover coarticulation between vowels and consonants, may influence categorization and discrimination of nonnative contrasts in different ways (see Avery and Best, 1995). Neuropsychological studies could also provide insights. To illustrate, in a recent dichotic listening study, although American English speakers and Zulu speakers displayed similar overall performance levels in judgments of Zulu click consonants, only the Zulu listeners perceived them as speech and showed a left hemisphere advantage (Best and Avery, 1999), indicating one crucial difference in perception of nonlinguistic versus phonological information in click consonants.

To sum up, the present findings are consistent with the hypothesis that non-native speech perception is based on detection of articulatory-phonetic similarities to the phonological units and contrasts of the native language. Discrimination performance levels are strongly linked to listeners' assimilations of non-native phones within their native phonological system. To a large extent, assimilation and discrimination of non-native consonants reflects listeners' sensitivity to phonetic and/or phonological similarities to native consonants. The detection of nonlinguistic properties in speech contributes minimally to non-native speech perception, being evident only when listeners reported hearing no phonetic or phonological differences between contrasting non-native consonants, in which case they showed fairly poor discrimination. The full set of results is most compatible with PAM predictions. While certain findings may be consistent with other views of non-native speech perception, PAM alone provided the motivation for the present cross-language comparisons, and it appears to offer the most coherent account.

Acknowledgments

References

Abramson, AS.; Lisker, L. Discriminability along the voicing continuum: Cross-language tests. Proceedings of the Sixth International Congress of Phonetic Sciences; Prague: Academia; 1970. p. 569-573.

Aslin, RN.; Pisoni, DB. Some developmental processes in speech perception. In: Yeni-Komshian, GH.; Kavanagh, JF.; Ferguson, CA., editors. Child Phonology: Perception. Academic; New York: 1980.

Avery RA, Best CT. Phonological and phonotactic influences on perception of two nonnative vowel contrasts. J Acoust Soc Am 1995;97:3362.

Best, CT. Phonetic influences on the perception of nonnativce speech contrasts by 6–8 and 10–12 month-olds. presented at the Society for Research in Child Development; Seattle. April; 1991.

Best, CT. Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In: de Boysson-Bardies, B.; de Schonen, S.; Jusczyk, P.; MacNeilage,

P.; Morton, J., editors. Developmental Neurocognition: Speech and Face Processing in the First Year. Kluwer Academic; Dordrecht, The Netherlands: 1993.

Best, CT. Learning to perceive the sound pattern of English. In: Rovee-Collier, C.; Lip-sitt, LP., editors. Advances in Infancy Research. Ablex; Norwood, NJ: 1994a.

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. MIT; Cambridge, MA: 1994b.

Best, CT. A direct realist perspective on cross-language speech perception. In: Strange, W., editor. Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-language Speech Research. York: Timonium, MD; 1995. p. 167-200.

Best CT, Avery RA. Left-hemisphere advantage for click consonants is determined by linguistic significance and experience. Psychological Science 1999;10:65–69.

Best CT, Strange W. Effects of phonological and phonetic factors on cross-language perception on approximants. J Phonetics 1992;20:305–330.

Best CT, Faber A, Levitt AG. Perceptual assimilation of non-native vowel contrasts to the American English vowel system. J Acoust Soc Am 1996;99:2602.

Best CT, McRoberts GW, Sithole NM. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. J Exp Psychol Hum Percept Perform 1988;4:45–60.

Best CT, Morrongiello B, Robson R. Perceptual equivalence of acoustic cues in speech and nonspeecch perception. Percept Psychophys 1981;29:191–211. [PubMed: 7267271]

Best CT, McRoberts GW, LaFleur R, Silver-Isenstadt J. Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. Infant Behavior and Development 1995;18:339–350.

Best, CT.; McRoberts, GW.; Goodell, E.; Womer, J.; Insabella, G.; Klatt, L.; Luke, S.; Silver, J. Infant and adult perception of nonnative speech contrasts differing in relation to the listeners' native phonology. presented at the International Conference in Infant Studies; Montreal. 19–22 April; 1990.

Browman CP, Goldstein L. Towards an articulatory phonology. Phonology Yearbook 1986;3:219–252.

Browman CP, Goldstein L. Articulatory gestures as phonological units. Phonology 1989;6:151–206.

Browman CP, Goldstein L. Gestural specification using dynamically-defined articulatory structures. J Phonetics 1990a;18:299–320.

Browman CP, Goldstein L. Representation and reality: physical systems and phonological structure. J Phonetics 1990b;18:411–424.

Browman CP, Goldstein L. Articulatory phonology: An overview. Phonetica 1992;49:155–180. [PubMed: 1488456]

Calderón J, Best CT. Effects of bilingualism on the perception of nonnative consonant voicing contrasts. J Acoust Soc Am 1996;99:2602.

Crowder RG. The sound of vowels and consonants in immediate memory. J Verbal Learn Verbal Behav 1971;10:587–596.

Crowder RG. Representation of speech sounds in precategorical acoustic storage. J Exp Psychol 1973;98:14–24. [PubMed: 4704206]

Diehl R, Kluender KR. On the objects of speech perception. Ecological Psychol 1989;1:1–45.

Dooling, RJ. Perception of complex, species-specific vocalizations by birds and humans. In: Dooling, RJ.; Hulse, SH., editors. The Comparative Psychology of Audition: Perceiving Complex Sounds. Erlbaum; Hillsdale, NJ: 1989.

Eimas PD. Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. Percept Psychophys 1975;18:341–347.

Eimas, PD. Comments: Some effects of language acquisition on speech perception. In: Mattingly, IG.; Studdert-Kennedy, M., editors. Modularity and the Motor Theory of Speech Perception. Erlbaum; Hillsdale, NJ: 1991.

Flege, JE. The production and perception of foreign language speech sounds. In: Winitz, H., editor. Human Communication and its Disorders. Vol. 2. Ablex; Norwood, NJ: 1986. p. 224-401.

Flege JE. Chinese subjects' perception of the word-final English /t/-/d/ contrast: Performance before and after training. J Acoust Soc Am 1989;86:1684–1697. [PubMed: 2808918]

Flege, JE. Perception and production: The relevance of phonetic input to L2 language learning. In: Ferguson, C.; Heubner, T., editors. Crosscurrents in Second Language Acquisition and Linguistic Theories. J Benjamin; Philadelphia: 1990.

Flege, JE. Second language speech learning: Theory, findings and problems. In: Strange, W., editor. Speech Perception and Linguistic Experience: Issues in Cross-language Speech Research. York: Timonium, MD; 1995. p. 233-272.

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

Fowler CA. Real objects of speech perception: A commentary on Diehl and Kluender. Ecological Psychol 1989;1:145–160.

Fowler CA, Best CT, McRoberts GW. Young infants' perception of liquid coarticulatory influences on following stop consonants. Percept Psychophys 1990;48(6):559–570. [PubMed: 2270188]

Francis, WN.; Kucera, H. Frequency analysis of English usage: Lexicon and grammar. Houghton-Mifflin; Boston: 1982.

Frieda EM, Walley AC, Flege JE, Sloane ME. Adults' perception of native and nonnative vowels: Implications for the perceptual magnet effect. Percept Psychophys 1999;61:561–577. [PubMed: 10334101]

Gibson JJ, Gibson EJ. Perceptual learning: Differentiation or enrichment? Psychol Rev 1955;62:32–41. [PubMed: 14357525]

Goto H. Auditory perception by normal Japanese adults of the sounds 'L' and 'R'. Neuropsychologia 1971;9:317–323. [PubMed: 5149302]

Grieser DL, Kuhl PK. Categorization of speech by infants: Support for speech-sound prototypes. Developmental Psychology 1989;25:577–588.

Guion SG, Flege JE, Akahane-Yamada R, Pruitt JC. An investigation of second language speech perception: The case of Japanese adults' perception of English consonants. J Acoust Soc Am 2000;107:2711–2724. [PubMed: 10830393]

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.

Hallé PA, Best CT, Levitt A. Phonetic versus phonological influences of French listeners' perception of American English approximants. J Phonetics 1999;27:281–306.

Hallé PA, Chéreau C, Segui J. Where is the /b/ in 'absurde'? It is in French listeners' minds. J Memory Language 2000;43:618–639.

Hallé PA, Segui J, Frauenfelder U, Meunier C. The processing of illegal consonant clusters: A case of perceptual assimilation? J Experimental Psychology: Human Perception and Performance 1998;24:1–17.

Iverson P, Kuhl PK. Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. J Acoust Soc Am 1996;99:1130–1140. [PubMed: 8609297]

Keller, E. Signalyze: Signal Analysis for Speech and Sound: Version 3.0. Infosignal, Inc; Lausanne, Switzerland: 1994.

Kluender, K. Speech perception as a tractable problem in cognitive science. In: Gernsbacher, MA., editor. Handbook of Psycholinguistics. Academic; San Diego, CA: 1994.

Kuhl PK. Auditory perception and the evolution of speech. Human Evolution 1988;3:19–43.

Kuhl PK. Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. Perception and Psychophysics 1991;50(2):93–107. [PubMed: 1945741]

Kuhl, PK. Speech prototypes: Studies on the nature, function, ontogeny and phylogeny of the "centers" of speech categories. In: Tohkura, Y.; Vatikiotis-Bateson, E.; Sagisaka, Y., editors. Speech Perception, Production and Linguistic Structure. Ohmsha; Tokyo: 1992. p. 239-264.

Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B. Linguistic experience alters phonetic perception in infants by 6 months of age. Science 1992;255:606–608. [PubMed: 1736364]

Ladefoged, P.; Maddieson, I. The Sounds of the World's Languages. Blackwell; Malden, MA: 1996.

Liberman AM, Mattingly IG. A specialization for speech perception. Science 1989;243:489–494. [PubMed: 2643163]

Liberman AM, Cooper FS, Shankweiler DS, Studdert-Kennedy M. Perception of the speech code. Psychological Review 1967;74:431–461. [PubMed: 4170865]

Lisker L, Abramson AS. The voicing dimension: Some experiments in comparative phonetics. Haskins Laboratories Status Report on Speech Research SR 1967;11:9–15.

Lively SE. An examination of the perceptual magnet effect. J Acoust Soc Am 1993;93:2423.

Lively SE, Logan JS, Pisoni DB. Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. J Acoust Soc Am 1993;94:1242–1255. [PubMed: 8408964]

Lively SE, Logan JS, Pisoni DB, Yamada RA, Tohkura Y, Yamada T. Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories. J Acoust Soc Am 1994;96:2076–2087. [PubMed: 7963022]

Logan JS, Lively SE, Pisoni DB. Training Japanese listeners to identify English /r/ and /l/: A first report. J Acoust Soc Am 1991;89:874–886. [PubMed: 2016438]

Lotto AJ, Kluender KR, Holt LL. Depolarizing the perceptual magnet effect. J Acoust Soc Am 1998;103:3648–3655. [PubMed: 9637046]

Lotto, AJ.; Kluender, KR.; Holt, LL. Effects of language experience on perceptual organization of vowel sounds. presented at the Fifth Conference on Laboratory Phonology; 6 July; Chicago: Northwestern University; 1996.

MacKain KS. Assessing the role of experience on infants' speech discrimination. J Child Language 1982;9:527–542.

MacKain KS, Best CT, Strange W. Categorical perception of English /r/ and /l/ by Japanese bilinguals. Applied Psycholinguistics 1981;2:369–390.

MacMillan, NA.; Creelman, CD. Detection Theory: A User's Guide. Cambridge U. P; Cambridge, England: 1991.

Maddieson, I. Patterns of Sounds. Cambridge U. P; Cambridge: 1984.

Miller GA, Nicely PE. Perceptual confusions among some English consonants. J Acoust Soc Am 1955;27:338–352.

Miller JL. On the internal structure of phonetic categories. Cognition 1994;50:271–285. [PubMed: 8039364]

Miyawaki K, Strange W, Verbrugge R, Liberman AM, Jenkins JJ, Fujimura O. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. Percept Psychophys 1975;18:331–340.

Mochizuki M. The identification of /r/ and /l/ in natural and synthesized speech. J Phonetics 1981;9:283–303.

Nittrouer S, Studdert-Kennedy M, McGowan RS. The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. J Speech Hear Res 1989;32:120–132. [PubMed: 2704187]

Pallier C, Bosch L, Sébastian-Galles N. A limit on behavioral plasticity in speech perception. Cognition 1997;64:9–17.

Pisoni DB, Aslin RN, Perey AJ, Hennessy BL. Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. J Exp Psychol 1982;8:297–314.

Polivanov E. La perception des sons d'une langue étrangère. Travaux du Cercle Linguistique de Prague 1931;4:79–96.

Polka L. Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. J Acoust Soc Am 1991;89:2961–2977. [PubMed: 1918634]

Polka L. Characterizing the influence of native experience on adult speech perception. Percept Psychophys 1992;52:37–52. [PubMed: 1635856]

Polka L, Bohn OS. A cross-language comparison of vowel perception in English-learning and German-learning infants. J Acoust Soc Am 1996;100:577–592. [PubMed: 8675849]

Polka L, Werker JF. Developmental changes in perception of nonnative vowel contrasts. J Experimental Psychology: Human Perception and Performance 1994;20:421–435.

Ruhlen, M. A Guide to the Languages of the World. Stanford Univ; Palo Alto, CA: 1975.

Scarborough HS. Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming, and iQ. Annals of Dyslexia 1998;48:115–136.

Shankweiler D, Crain S, Katz L, Fowler AE, Liberman AM, Brady SA, Thornton R, Lundquist E, Dreyer L, Fletcher JM, Stuebing KK, Shaywitz SE, Shaywitz BA. Cognitive profiles of reading-disabled children: Comparison of language skills in phonology, morphology, and syntax. Psychological Science 1995;6:149–156.

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

Strange W, Dittmann S. Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. Percept Psychophys 1984;36:131–145. [PubMed: 6514522]

Sussman JE, Lauckner-Morano VJ. Further tests of the "perceptual magnet effect". in the perception of [i]: Identification and change-no-change discrimination. J Acout Soc Am 1995;97:539–552.

Takagi N, Mann VA. Signal detection modeling of Japanese listeners' /r/-/l/ labeling behavior in a one-interval identification task. J Acoust Soc Am 1995;97:563–574. [PubMed: 7860833]

Tees RC, Werker JF. Perceptual flexibility: Maintenance or recovery of the ability to discriminate nonnative speech sounds. Can J Psychol 1984;38:579–590. [PubMed: 6518419]

Trubetzkoy, NS. Principles of Phonology. Baltaxe, CAM., editor. Univ. California; Berkeley, CA: 19391969.

Volaitis LE, Miller JL. Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. J Acoust Soc Am 1992;92:723–735. [PubMed: 1506527]

Werker JF. On becoming a native listener. Am Sci 1989;77:54–59.

Werker JF, Lalonde CE. Cross-language speech perception: Initial capabilities and developmental change. Developmental Psychology 1989;24:672–683.

Werker JF, Logan J. Cross-language evidence for three factors in speech perception. Percept Psychophys 1985;37:35–44. [PubMed: 3991316]

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

Werker JF, Tees RC. Phonemic and phonetic factors in adult cross-language speech perception. J Acoust Soc Am 1984;75:1866–1878. [PubMed: 6747097]

Werker JF, Gilbert JHV, Humphrey K, Tees RC. Developmental aspects of cross-language speech perception. Child Development 1981;52:349–355. [PubMed: 7238150]

Whalen DH. Subcategorical phonetic mismatches slow phonetic judgments. Percept Psychophys 1984;35:49–64. [PubMed: 6709474]

Whalen DH. Subcategorical phonetic mismatches and lexical access. Percept Psychophys 1991;50:351–360. [PubMed: 1758766]

Whalen DH, Wiley ER, Rubin PE, Cooper FS. The Haskins Laboratories' pulse code modulation (PCM) system. Behav Res Methods Instrum Comput 1990;22:550–559.

Yamada R, Tohkura Y. The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. Percept Psychophys 1992;52:376–392. [PubMed: 1437471]
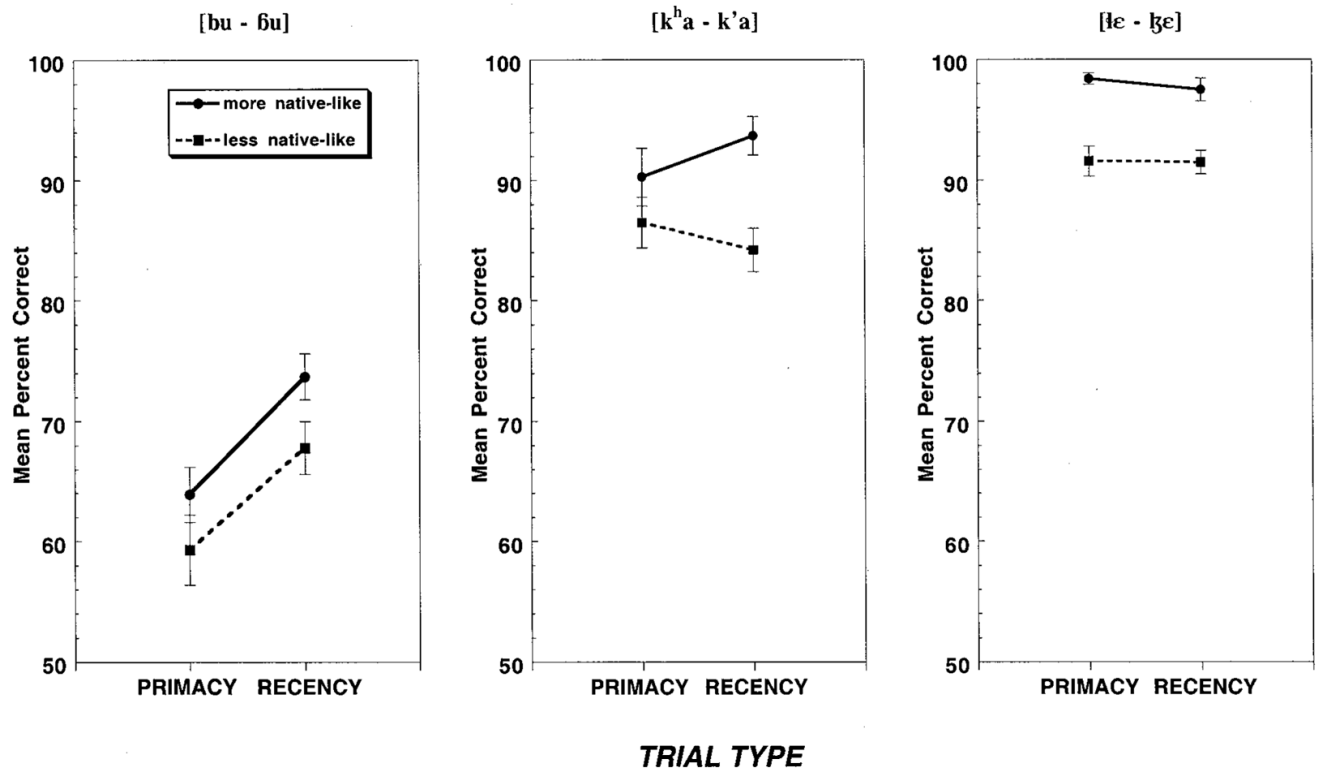
FIG. 1.
The AXB discrimination performance in experiment 1 for the factors of contrast×trial type×native likeness. The three panels display results for (a) Zulu plosive versus implosive bilabial stops (SC), (b) voiceless aspirated versus ejective velar stops (CG), and (c) voiceless versus voiced lateral fricatives (TC).

TABLE I

Acoustic attributes of the stimulus tokens selected for each target category in experiment 1.

Stimulus syllable	Syllable duration[a]	Consonant duration[a]	Vowel duration[a]	Consonant VOT[a]	Consonant amplitude[b]	Consonant centroid[c]	Vowel F1[d]	Vowel F2[d]	Vowel F3[d]	Vowel F0[d]
Lateral fricatives										
/ʒe/ (voiced)										
first pulse[f]	**309.9** (271–336)[e]	**108.3** (86–131)	**197.6** (185–217)		**25.3** (22–28)		**475.2** (436–499)	**2318.9** (2259–2384)	**2813.4** (2623–2951)	**204.8** (153–233)
15%[g]						3512.3 (3239–3722)	528.0 (499–592)	2443.6 (2384–2586)	2812.1 (2638–2943)	221.5 (214–232)
50%[g]						3616.5 (3549–3767)	563.0 (554–608)	2560.4 (2493–2586)	2875.5 (2676–2996)	224.0 (215–233)
85%[g]						3683.5 (3572–3805)	550.5 (514–592)	2539.7 (2477–2586)	2873.0 (2780–2973)	226.0 (220–231)
/ɬe/ (voiceless)										
first pulse	**345.2** (299–400)	**151.7** (134–179)	**187.9** (161–216)		**31.8** (24–36)		**542.7** (530–577)	**2279.5** (2228–2322)	**2850.6** (2810–2899)	**246.2** (212–267)
15%						3654.9 (3621–3685)	579.1 (561–592)	2374.1 (2322–2430)	2800.9 (2705–2907)	240.4 (236–244)
50%						3704.2 (3636–3810)	592.1 (577–608)	2495.5 (2337–2571)	2818.3 (2690–2936)	232.5 (227–239)
85%						3690.3 (3550–3810)	555.7 (514–577)	2557.8 (2509–2633)	2900.3 (2847–2959)	229.8 (221–238)
Velar stops										
/kʰa/ (aspirated)										
first pitch pulse	**284.8** (246–314)	**8.5** (7–11)	**202.8** (158–234)	**82.5** (76–88)	**30.2** (28–33)		**1075.7** (1006–1162)	**1546.4** (1468–1655)	**2439.5** (2325–2534)	**225.2** (201–262)
15%						2059.1 (1841–2423)	1076.9 (991–1177)	1519.1 (1431–1610)	2505.3 (2392–2646)	209.7 (201–219)
50%						1972.1 (1876–2268)	1063.2 (1014–1125)	1442.1 (1379–1535)	2523.9 (2467–2673)	197.6 (185–204)
85%						1913.7 (1814–2164)	1049.6 (1006–1095)	1525.3 (1438–1714)	2632.0 (2556–2750)	195.5 (188–203)
/kʼa/ (ejective)										
first pitch puls	**263.5** (231–278)	**37.3** (23–55)	**175.4** (154–194)	**88.3** (70–113)	**42.3** (38–47)		**1063.3** (1029–1117)	**1455.8** (1319–1528)	**2514.0** (2400–2661)	**197.1** (134–233)
15%						2580.3 (2391–2885)	1104.2 (1029–1178)	1604.8 (1528–1729)	2552.5 (2467–2623)	203.0 (185–213)
50%						2616.5 (2408–2788)	1099.3 (1050–1170)	1584.9 (1431–1729)	2624.6 (2430–2772)	198.3 (189–204)
85%						2709.3 (2477–2937)	1112.9 (1038–1200)	1578.7 (1498–1692)	2613.4 (2452–2743)	194.9 (189–200)
Bilabial stops										
/bu/ (plosive)										
first pitch pulse	**261.4** (232–278)	**9.7** (6–13)	**248.3** (211–267)	**13.5** (10–23)	**15.8** (13–17)		**463.4** (443–499)	**895.0** (836–919)	**2461.7** (2337–2586)	**227.5** (211–242)
15%						2806.8 (2284–3096)	475.1 (443–494)	935.1 (886–967)	2365.5 (2307–2448)	230.5 (213–239)
50%						2707.0 (2400–3220)	476.8 (453–494)	935.1 (866–987)	2387.4 (2317–2569)	228.5 (213–244)
85%						2758.1 (2378–3216)	478.5 (463–494)	955.3 (917–987)	2424.3 (2226–2851)	232.0 (225–239)
/ɓu/ (implosive)										
first pitch pulse	**293.8** (260–341)	**12.1** (7–33)	**221.5** (204–252)	**−58.7** (−105/−67)	**26.7** (23–31)		**450.2** (405–483)	**907.1** (873–927)	**2472.2** (2307–2680)	**230.1** (219–255)
15%						2637.8 (2389–2907)	529.8 (499–577)	877.7 (810–982)	2383.9 (2290–2493)	254.2 (238–270)
50%						2591.5 (2250–3022)	496.1 (483–514)	981.9 (935–1044)	2391.7 (2275–2462)	246.4 (227–256)
85%						2287.5 (1982–2782)	493.4 (483–499)	937.5 (826–997)	2370.9 (2275–2462)	237.8 (219–254)

[a] Duration in milliseconds (ms); consonant duration refers to burst for velar and bilabial stops, to frication for lateral fricatives; VOT=voice onset time for stops only.

[b] Root mean square amplitude across full duration of consonant noise (frication or burst).

[c] Centroid frequencies of consonant noise at onset, middle, and offset of noise (frication or burst).

[d] Frequency in Hz at first pitch pulse and at 15%, 50%, and 85% into vowel for lowest three formants and F0 (LPC estimates).

[e] Minimum and maximum values, to nearest integer.

[f] First pitch pulse of vocalic portion.

[g] Percent into consonant noise (frication or burst) for centroid measures; percent into vocalic portion for formant and F0 measures.

TABLE II

Mean percent correct discrimination for trial type×native likeness×contrast.

| | Trial type | | | |
| | Primacy (AAB, BBA) | | Recency (BAA, ABB) | |
Contrast	More nativelike	Less nativelike	More nativelike	Less nativelike
Zulu [bu–ɓu]	63.98 (27.97)[a]	59.32 (18.66)	73.67 (47.34)	67.85 (35.69)
Zulu [kʰa–k'a]	90.35 (80.71)	86.45 (72.90)	93.68 (87.36)	84.17 (68.34)
Zulu [ɮɛ–ɬɛ]	98.42 (96.86)	91.57 (83.14)	97.52 (95.04)	91.46 (83.29)

[a] Values in ( ) are corrected for guessing/bias (Macmillan and Creelman, 1991; see text), i.e., corrected percent above chance.

TABLE III

American English-speaking adults' assimilations of Zulu consonants. Values in parentheses indicate number of participants providing each transcription type.

	Lateral fricatives		Velar stops		Bilabial stops	
	Voiceless: /ɬe/	**Voiced: /ɮe/**	**Voiceless: /ka/ ([kʰal])**	**Ejective: /k'a/**	**Plosive: /bu/**	**Implosive: /ɓu/**
	(6) s[a]	(8) l[b]	(19) k[c,d]	(10) k[c]	(16) b	(20) b[e]
	(8) sh	(3) z	(2) ḵ[c,f]	(5) kh	(4) bh	(1) v
	(4) sl[g]	(3) zl[h]	(1) ck[i]	(2) chz	(1) missed	(1) vb
	(2) ch	(1) thl	kh[j]	(2)	(1)	mb
	(2) ts[k]	(6) th with z/v	kch[l]	(2)		
	cth	(1) szt		(1)		

[a] Includes "s" or "hs" or "soft' c" or "'soft' sc."

[b] Includes "hl."

[c] Includes "k" or "'hard' c."

[d] Described by all listeners with additional features: choking, throat-clearing, clacking, clicking, gagging.

[e] Described by four listeners as "harder" or "with pursed lips" or "with tensed speech muscles."

[f] Includes transcriptions of "k" with apostrophe or with epenthetic vowel after "k."

[g] Includes "shl" or "chl."

[h] Includes "zhl."

[i] Includes "qk" or "gk" or "tk."

[j] Includes "gh."

[k] Includes "tz" or "tsc" or "zs."

[l] Includes "chg."

TABLE IV

Acoustic attributes of the stimulus tokens selected for each target category in experiment 2.

Stimulus syllable	Syllable duration[a]	Consonant duration[a]	Vowel duration[a]	Consonant VOT[a]	Consonant amplitude[b]	Consonant centroid[c]	Vowel F1[d]	Vowel F2[d]	Vowel F3[d]	Vowel F0[d]
Tigrinya ejectives:										
/p'ɛ/[c]	235.0 (223–248)[c]									
first pulse[f]		12.0 (9–16)	157.9 (144–169)	74.8 (66–92)	40.8 (37–43)		456.7 (318–507)	2001.5 (1871–2190)	2632.0 (2517–2780)	127.1 (104–143)
15%[g]						3602.3 (3282–3993)	618.7 (593–627)	2052.3 (2027–2086)	2868.0 (2643–2998)	107.3 (105–111)
50%[g]						3365.8 (2977–3780)	626.3 (612–637)	2085.8 (2023–2205)	2792.8 (2742–2815)	97.3 (94–101)
85%[g]						3223.5 (2991–3595)	624.0 (612–630)	2068.5 (2028–2209)	2872.7 (2815–2998)	90.7 (86–97)
/t'ɛ/	232.8 (213–248)									
first pulse		20.6 (18–27)	157.1 (135–176)	75.1 (57–105)	48.7 (47–52)		478.0 (472–490)	2517.5 (2448–2570)	3814.2 (3724–3986)	126.7 (111–145)
15%						4614.0 (4391–4848)	622.2 (612–631)	2078.0 (1885–2195)	2863.8 (2805–2937)	114.9 (106–123)
50%						4494.2 (3958–5112)	621.3 (612–631)	2088.0 (2033–2189)	2882.5 (2804–3015)	102.2 (93–106)
85%						4042.4 (3372–4874)	611.2 (551–630)	2142.3 (2033–2199)	2818.5 (2805–2936)	91.6 (86–101)
Eng. alveolar fricatives										
/sɛ/	487.8 (471–509)	239.3 (221–254)	319.3 (313–329)		50.3 (49–52)					
first pitch pulse							319.3 (313–329)	1677.2 (1557–1741)	2785.2 (2645–2817)	218.3 (186–237)
15%						6822.3 (6511–6971)	477.2 (468–487)	1731.3 (1713–1756)	2693.5 (2505–2821)	188.6 (188–192)
50%						7177.2 (6992–7473)	486.7 (475–496)	1722.5 (1575–1878)	2711.5 (2639–2822)	172.8 (165–174)
85%						7348.0 (6965–7750)	476.2 (461–496)	1648.2 (1617–1692)	2617.0 (2508–2791)	166.0 (159–178)
/zɛ/	465.0 (440–497)	211.8 (197–244)	238.6 (222–253)		48.8 (47–51)					
first pitch pulse							308.5 (299–319)	1720.3 (1554–1881)	2725.8 (2517–2804)	185.4 (179–189)
15%						6345.8 (5533–7182)	477.0 (468–483)	1781.3 (1713–1881)	2810.2 (2801–2821)	184.8 (180–190)
50%						7177.6 (6809–7563)	482.3 (466–498)	1731.3 (1554–1876)	2765.3 (2672–2831)	172.6 (168–177)
85%						7402.2 (7175–7768)	466.5 (455–481)	1698.5 (1564–1881)	2737.5 (2500–2819)	158.1 (153–161)
Eng. palatal fricatives										
/ʃɛ/	490.7 (476–500)	256.7 (243–268)	229.1 (223–237)		46.5 (45–48)					
first pitch pulse							325.0 (313–348)	1618.5 (1541–1866)	2474.0 (2203–2657)	227.5 (215–239)
15%						5477.5 (5250–5644)	478.2 (458–498)	1853.5 (1741–1866)	2740.2 (2517–2821)	186.4 (180–195)
50%						5629.3 (5473–5818)	483.0 (473–495)	1801.3 (1716–1886)	2630.7 (2495–2786)	171.4 (165–180)
85%						5539.3 (5451–5613)	482.5 (472–490)	1732.1 (1724–1744)	2714.3 (2508–3129)	168.0 (164–174)
/ʒɛ/	485.6 (467–509)	229.5 (213–256)	242.9 (227–264)		45.3 (43–47)					
first pitch pulse							316.4 (313–319)	1650.3 (1556–1783)	2522.0 (2360–2640)	184.5 (175–193)
15%						5134.6 (4912–5619)	479.7 (461–496)	1879.2 (1867–1885)	2663.0 (2500–2813)	179.8 (174–185)
50%						5477.7 (5290–5924)	479.0 (464–490)	1795.5 (1709–1876)	2672.3 (2477–2809)	170.9 (167–177)
85%						5489.3 (5278–5740)	482.0 (471–489)	1708.8 (1570–1766)	2631.2 (2454–2815)	169.3 (166–172)

[a] Duration in milliseconds (ms); consonant duration refers to burst for velar and bilabial stops, to frication for lateral fricatives.

[b] Root mean square amplitude across full duration of consonant noise (frication or burst).

[c] Centroid frequencies of consonant noise at onset, middle, and offset of noise (frication or burst).

[d] Frequency in Hz at first pitch pulse and at 15%, 50%, and 85% into vowel for lowest three formants and F0 (LPC estimates).

[e] Minimum and maximum values, to nearest integer.

[f] First pitch pulse of vocalic portion.

[g] Percent into consonant noise (frication or burst) for centroid measures; percent into vocalic portion for formant and F0 measures.

TABLE V

Mean percent correct discrimination for trial type×contrast in experiment 2.

| Contrast | Trial type | | | |
| | Primacy | | Recency | |
	AAB	BBA	BAA	ABB
English [sɛ–zɛ]	97.66 (95.32)[a]	99.42 (98.83)	99.71 (99.42)	98.25 (96.49)
English [ʃɛ–ʒɛ]	98.24 (96.49)	98.54 (97.08)	99.71 (99.42)	98.54 (97.08)
Tigrinya [p′ ɛ–t′ ɛ]	91.81 (83.62)	88.59 (77.19)	92.11 (84.21)	92.98 (85.96)

[a]Values in ( ) are corrected for guessing/bias (Macmillan and Creelman, 1991; see text), i.e., corrected percent above chance.