638 3

Journal of Experimental Psychology:
Human Perception and Performance
1988, Vol. 14, No. 3, 345-360

Examination of Perceptual Reorganization for Nonnative Speech Contrasts: Zulu Click Discrimination by English-Speaking Adults and Infants

Catherine T. Best
Wesleyan University
and Haskins Latoratories
New Haven, Connecticut

Gerald W. McRoberts
University of Connecticut
and Haskins Laboratories,
New Haven, Connecticut

Nomathemba M. Sithole
Wesleyan University

The language environment modifies the speech perception abilities found in early development. In particular, adults have difficulty perceiving many nonnative contrasts that young infants discriminate. The underlying perceptual reorganization apparently occurs by 10–12 months. According to one view, it depends on experiential effects on psychoacoustic mechanisms. Alternatively, phonological development has been held responsible, with perception influenced by whether the nonnative sounds occur allophonically in the native language. We hypothesized that a phonemic process appears around 10–12 months that assimilates speech sounds to native categories whenever possible; otherwise, they are perceived in auditory or phonetic (articulatory) terms. We tested this with English-speaking listeners by using Zulu click contrasts. Adults discriminated the click contrasts; performance on the most difficult (80% correct) was not diminished even when the most obvious acoustic difference was eliminated. Infants showed good discrimination of the acoustically modified contrast even by 12–14 months. Together with earlier reports of developmental change in perception of nonnative contrasts, these findings support a phonological explanation of language-specific reorganization in speech perception.

Infants in the first half-year or so of life discriminate most speech sound distinctions with which they have been tested, including sounds that are not contrasted phonologically in the language of their environment, that is, are not used to specify differences in word meanings, such as the English /r/–/l/ contrast that is irrelevant in Japanese (Aslin, Pisoni, Hennessy, & Perey, 1981; Best, 1984; Jusczyk, 1984; Lasky, Syrdal-Lasky, & Klein, 1975; Trehub, 1976; Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 1984a). In some cases, the language environment may instead facilitate infants' initially poor discrimination of certain other nonnative contrasts, such as English /s/-/z/ (e.g., Aslin & Pisoni, 1980; Eilers, Gavin, & Oller, 1982; Eilers, Gavin, & Wilson, 1979; Eilers & Minifie, 1975; Eilers, Wilson, & Moore, 1977; Streeter, 1976). Adults, on the other hand, typically discriminate all native contrasts but have difficulty discriminating nonnative contrasts (e.g., Abramson & Lisker, 1970; Goto, 1971; MacKain, Best, & Strange, 1981; Miyawaki et al., 1975; Singh & Black, 1966; Tees & Werker, 1984; Trehub, 1976; Werker et al., 1981; Werker & Logan, 1985; Werker & Tees, 1984b). The present report is concerned in particular with understanding developmental change in perception of early-discriminated nonnative speech contrasts.

The language environment clearly influences developmental speech perception. But is the influence due simply to differential *auditory* exposure, or does it derive instead from the *linguistic* experience of acquiring and using the phonological system of the native language? In line with the latter suggestion, Eimas (1978) suggested that the speech perception abilities of infancy become reorganized by adulthood as a function of specific *linguistic* experience. Werker and her colleagues demonstrated that the perceptual reorganization has already occurred by 4 years of age (Werker & Tees, 1983) and, in fact, appears to take place around 10–12 months of age (Werker et al., 1981; Werker & Tees, 1984a; but see also Burnham, 1986). They also proposed a linguistic account of the reorganization—that infants presumably shift developmentally from perception of speech contrasts in prephonological terms, that is, based on their *acoustic* properties (phys-

---

This research was supported by National Institute of Health Grants HD-01994 and RR-05596 to Haskins Laboratories and by National Institute of Health Grant NS-24655 and a Biomedical Research Support grant awarded through Wesleyan University to Catherine T. Best.

We wish to thank the following people for their invaluable help in completing this research: Janet Werker for numerous useful discussions about speech perceptual development, helpful methodological advice, and the loan of her /ba/–/da/ stimuli; Andrea Levitt, Linda Polka, Bruno Repp, Michael Studdert-Kennedy, and three reviewers for thoughtful comments on earlier versions of the article; Tshitshi Mbatha and Otty Nxumalo for producing the Zulu click syllables from which our stimuli were chosen; Suzanne Margiano, Cynthia Nye, and Dianne Schrage for their assistance in collecting data from the infant subjects; our American and South African adult subjects; and most especially the parents who brought their infants to the laboratory to participate in the study.

Correspondence concerning this article should be addressed to Catherine T. Best, Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511.

ical characteristics such as frequency components, silent gaps, and noise bursts) and *phonetic* properties (characteristics of the way the sounds are articulated), to perception of them in terms of contrasts that occur in the phonological system of their language (see also Jusczyk, 1982; MacKain, 1982), referred to as *phoneme contrasts*. Such a developmental change would seem well suited to the infant's first steps in receptive and productive acquisition of words near the end of the first year (e.g., Lenneberg, 1967; Stark, 1980).

An alternative proposal, however, appeals to an auditory or psychoacoustic influence of experience, by which exposure to and processing of the acoustic properties of native speech sounds causes some change in the responsiveness or "tuning" of the listener's auditory system. The psychoacoustic proposal states that discrimination of early-distinguished nonnative contrasts declines because of lack of auditory experience with sounds that do not appear in native phoneme contrasts (e.g., Aslin & Pisoni, 1980). We refer to this as the *auditory experience* argument. Note that in either the psychoacoustic or the linguistic hypothesis, the perceptual change need not be permanent and may instead involve shifts in attentional mechanisms (e.g., Aslin & Pisoni, 1980; MacKain et al., 1981).

Any explanation of language-specific effects on speech perception should take into account the relation of phonetic properties to phonemic contrasts, which are defined by particular combinations of phonetic features. To illustrate, /b/–/p/ (represented at the phonemic level) share the phonetic features of bilabial closure and stop or obstruent manner of production, but they differ in voicing. The phones (phonetic representations that don't specify their phonological status in a given language) in this contrast, in utterance-initial position in English as in <bat>–<pat>, are [b]–[pʰ] or [p]–[pʰ]. That is, the phoneme /b/ typically has the phonetic feature either of voicing that occurs slightly before or simultaneous with the bilabial release ([b]) or of a short delay in voicing ([p]), and the phoneme /p/ has the phonetic feature of aspiration ([pʰ]) during the longer delay of voicing (Goldstein & Browman, 1986). MacKain (1982) and Tees and Werker (1984) point out that auditory experience with a given pair of phones is not ruled out simply because they fail to appear in a native phoneme contrast. Phone differences that are put to phonological use in some foreign language, but not in the native language, may nonetheless occur as allophones, or within-category phonetic variations, of some native category (although many phones that appear in nonnative phoneme contrasts never, or only very rarely, occur as allophones of native phonemes). In those cases of allophonic representation, auditory experience would occur without corresponding to phonological relevance. For example, some Arabic languages such as Farsi include a phonological contrast between the voiced velar stop /g/, and a voiced uvular stop /G/ that does not contrast with /g/ in English (Maddieson, 1984). Yet certain cases of the phone [G] can occur in English as allophonic variants of the phoneme /g/ in the context of back vowels such as /u/ and /a/, which cause /g/ to be articulated farther back than in the context of front vowels such as /i/.

The linguistic proposal has thus been extended to address nonnative contrasts that occur as allophonic variants in the native language. Werker et al. (1981; Werker and Tees, 1984a) argued that developmental decline in discrimination occurs for contrasting sounds that do not partake in native phonological contrasts, even if they occur as native allophonic variants. We refer to this as the *specific phonological relevance* hypothesis. A problem for this hypothesis, as well as for the auditory experience hypothesis, is that there appears to be some variability in the degree of perceptual reorganization for various nonnative contrasts. Recent findings reveal relatively good discrimination of nonnative Hindi aspirated voiced versus voiceless stops /dʰ/–/tʰ/ (Werker & Tees, 1984b) and Farsi /g/–/G/ by English-speaking adults (Polka, 1987), and of nonnative /w/–/r/ by Japanese-speaking adults (Best, MacKain, & Strange, 1982).

To address this problem, Tees and Werker (1984) proposed that allophonic variants may provide the listener with experience that maintains some discrimination of phonetically similar nonnative phones, an *allophonic experience* hypothesis. Indeed, English-speaking adults are able to discriminate, especially after perceptual training or reduced memory load, nonnative contrasts in which the members occur as allophonic variants in English, such as Hindi [dʰ]–[tʰ] and the utterance-initial prevoiced versus voiceless unaspirated [b]–[p] found in Spanish and other languages (Tees & Werker, 1984; Werker & Logan, 1985; Werker & Tees, 1984b; see also Carney, Widin, & Viemeister, 1977; Pisoni, Aslin, Perey, & Hennessy, 1982; Pisoni & Lazarus, 1974). In contrast, listeners have persistent difficulty with many nonnative distinctions in which one or both members fail to appear allophonically in English (Tees & Werker, 1984). However, this explanation does not distinguish between nonnative contrasts whose members are allophonic variants of a single native phonemic category versus those that may be variants of different native categories. The Farsi /g/–/G/ contrast is an example of the former case for English-speaking listeners. As an example of the latter, in intervocalic position the apical flap [ɾ] and the alveolar trill [r] are contrasted phonologically in Spanish but not English; however, [ɾ] is an allophonic variant of American English /t/ and /d/, and [r] is a British variant (Scottish accent) of American English [ɹ]. Note, though, that [ɾ]–[r] are not contrasted phonologically either in American English, which does not use the trill as an allophone of /r/, or in British English, which employs both [ɾ] and [r] as allophonic variants of /r/. The difference in allophonic status of /g/–/G/ versus /ɾ/–/r/ should lead to better discrimination of the latter than the former contrast by English-speaking listeners, if variability for nonnative contrasts does indeed depend on experience with allophones. Yet just the opposite pattern has been found (Oller & Eilers, 1983; Polka, 1987). Moreover, anecdotal evidence from English listeners suggests that certain other nonnative contrasts are quite discriminable even though neither element occurs as an allophonic variant in English, such as the glottalized stop distinction /k'/–/t'/ found in Tigrinya and !Xoo (Maddieson, 1984). However, other very similar contrasts that also fail to appear allophonically, such as the glottalized velar–uvular stop distinction /k'/–/q'/ which is found in Thompson (a Northwest American Indian language), are very difficult for English listeners to distinguish (Werker & Tees, 1984a, 1984b).

A psychoacoustic explanation of perceptual variability, that is, one that appeals to generalized rather than linguistically specialized responses of the auditory system to various acoustic properties, might seem to handle such exceptions more easily. Burnham (1986) suggested that relatively discriminable nonnative contrasts are distinguished from poorly discriminated ones because the former are psychoacoustically "robust" while the latter are psychoacoustically "fragile." Fragile contrasts are believed to be lost in infancy and remain difficult for adults even after perceptual training or reduction of memory demands. Robust contrasts, on the other hand, are presumably discriminated until at least 4–6 years and are more amenable to perceptual training in adulthood. Although this psychoacoustic approach might account for discriminability of nonnative contrasts that do not occur allophonically, it also has shortcomings. Most important, to avoid tautology, the hypothesis would require the establishment of language-independent criteria, such as a description of the acoustic features that should be associated with either end of the robust/fragile dimension, or a hierarchy of the relative difficulty that some nonhuman species has with various speech contrasts.

In the present research, we hypothesized that for listeners who have acquired the phonological system of their native language (or have begun to do so), attention is focused, during speech perception, predominantly at the phonemic level. For simplicity's sake, we refer to this as *phonemic perception* (see also Werker & Logan, 1985). It entails the perceptual assimilation[1] of incoming speech sounds to the phonemic categories of the native language whenever possible. Assimilation may take place regardless of whether those sounds are native or nonnative and regardless of whether they actually occur allophonically or are simply phonetically similar to some native category. Nonnative contrasts can be divided into four classes: those in which (a) the contrasting phones are assimilated as variants of a single native category ("single-category" assimilation); (b) the phones are assimilated as the opposing members of a native phonological contrast ("opposing-category" assimilation); (c) one member is better assimilated to a native category (more similar phonetically) than the other ("category-goodness difference" assimilation); and (d) both members are phonetically dissimilar from any native categories and are therefore not assimilated ("nonassimilation").

In single-category assimilations, discrimination should be difficult for adults, even with perceptual training, and decline in discrimination should occur by 10–12 months of age. The last three classes should be discriminated by adults and older infants, but for different reasons. Opposing-category assimilations should be perceived as phonemic contrasts. Category-goodness difference assimilations should be perceived as a difference in "goodness of fit" for a native phoneme category. That is, although attention is primarily focused at the phonemic level, listeners should retain some sensitivity to within-category phonetic articulatory variations that show differences in degree of match with the phonetic properties of the "ideal" category exemplar. Thus, the second and third classes, along with the first, involve perception of phonemically relevant information. In contradistinction, nonassimilated contrasts should be perceived in terms of their auditory (acoustic or nonspeech) or phonetic (phonologically neutral articulatory) characteristics.

Because of linguistic constraints on possible phonological oppositions, most of the contrasts of the world's languages naturally fall into the first three classes. Previous research, including Werker's studies of infants, has focused primarily on nonnative contrasts of the first class. Examples of single-category contrasts are Thompson /k'/–/q'/, which assimilate to English /k/ (Werker & Logan, 1985), and Spanish intervocalic /r/–/ɾ/ (Oller & Eilers, 1983), which should assimilate to /t/ or /d/ for American English listeners and to /r/ for British listeners. Perceptual difficulty with the latter contrast is problematic for the allophonic experience hypothesis. The second and third class are represented in studies that found relatively good discrimination, and effectiveness of perceptual training, in adults for some nonnative contrasts (e.g., Aslin et al., 1981; Best et al., 1982; Polka, 1987; Tees & Werker, 1984; Werker & Tees, 1984b). An example of an opposing-category contrast is /k'/–/t'/, which should be assimilable to English /k/–/t/. This presumably discriminable contrast would be troublesome not only for the allophonic and auditory experience hypotheses, because glottalized stops don't occur phonologically or allophonically in American English, but also for the robust–fragile psychoacoustic distinction, because the acoustic difference between the /k'/–/t'/ release bursts is likely analogous in magnitude to that found in the poorly discriminated Thompson /k'/–/q'/. Category-goodness difference assimilation is represented by the /g/–/G/ contrast that English speakers discriminate even though English does not contrast /g/–/G/ in any vowel context[2], and the English /w/–/r/ contrast that Japanese discriminate even though [ɹ] does not occur in Japanese. The listeners in the former category-goodness example reported hearing a good English /g/ versus a foreign-sounding one. In the latter, the Japanese subjects recognized /w/ as a good example of their native /w/, while /r/ was deviant from any native category. Neither example corresponds well to predictions of the auditory experience or the specific phonological relevance hypotheses.

The research reported here focused on nonassimilable nonnative contrasts. Specifically, we assessed whether English-speaking adults would show good discrimination for nonnative contrasts whose phonetic characteristics are highly dissimilar from any native categories and that are therefore unlikely

---

[1] The perceptual process described here may be similar to the notion of "phonetic analogy" mentioned by Eilers et al. (1982) and also to the concept of "phonic interference" that has been used to describe the spoken errors made by learners of a second language (e.g., Weinreich, 1953). Note that the process proposed here is not the same phenomenon as "phonemic assimilation" in speech production whereby, for example, <pocke*t*book> is pronounced as though it were <pocke*p*book>.

[2] It should be noted that changes in the phoneme context of a given phoneme (i.e., changes in surrounding vowels and/or consonants) may affect the relation of specific phonetic properties to the phoneme category of interest. For example, /g/ is produced farther back in the context of /u/ than of /i/. Such context effects do have perceptual ramifications (e.g., Mann & Repp, 1980).

to be assimilated. We also tested whether discrimination of these presumably nonassimilable contrasts depends on auditory cues that might be assumed to have robust psychoacoustic effects or rather on other auditory or phonetic (articulatory) differences that might be considered to be psychoacoustically fragile because of their similarity to certain perceptually difficult, early-lost nonnative contrasts. Finally, we tested whether infants show perceptual change at 10–12 months for such a contrast, as they do for single-category contrasts.

For this research, we used the click consonants of Zulu, which appear neither as phonemic contrasts nor as allophonic variants in English, nor do their phonetic–articulatory features correspond well to English phonemes. Although American listeners have typically experienced clicks produced as *non*speech "mouth sounds" or affectively toned vocalizations,[3] this does not constitute linguistic experience.[4] The clicks in spoken Zulu are produced with a vowel context and carry coarticulatory information as well as consonantal phonetic features such as voicing, nasalization, or glottalization. Nonspeech clicks have none of these phonetic characteristics.

In Experiment 1, we predicted that American English-speaking adults would be well able to discriminate click syllable contrasts, because they should not assimilate the clicks to English phonemes.

Experiment 1

Zulu, a Bantu language, is one of a number of tone languages from southern Africa that employ click consonants, which are ingressive, unlike any English consonants (Muller, 1965). Clicks are produced by the formation of a suction chamber in the oral cavity followed by an abrupt release of the negative pressure (Catford & Ladefoged, 1968; Doke, 1926; Ladefoged, 1971, 1975) at the blade, tip, or side of the tongue or at the lips (kissing sound) as in !Xoo, a Khoisan language (Ladefoged & Traill, 1984). Because the suction involves velar occlusion of airflow, click release in Zulu also includes subsequent velar release at varying delays of voicing (Doke, 1926). Zulu has 15 clicks, distributed across three different places of articulation: apicodental ([ʇ]), palatoalveolar ([ʗ]), and lateral alveolar ([ʖ]). Each is produced with one of five additional phonetic features. There are two categories of nasalized clicks (voiced or voiceless unaspirated) and three nonnasalized voicing categories (voiced, voiceless unaspirated, and voiceless aspirated) (Catford & Ladefoged, 1968; Doke, 1926; Maddieson, 1984; Nyembezi, 1972). We could find no published reports on perceptual or acoustic studies of the Zulu clicks, although Doke (1926) has described articulatory properties of Zulu clicks, and Ladefoged and Traill (1984) have described the articulatory and acoustic properties of clicks in several Khoisan languages. Zulu has a moderate number of clicks as compared with !Xoo, which has five places of articulation and 16 possible phonetic accompaniments (e.g., voicing, nasalization, glottalization, velarization, or combinations thereof) that can be applied at each place (Ladefoged & Traill, 1984).

In the Zulu apical (apicodental) click, the tongue tip is released from the back of the upper front teeth. For the palatal (palatoalveolar) click, the tongue tip and blade are released in midline at the front of the hard palate, behind the alveolar ridge. The lateral (lateral alveolar) click is asymmetrical, with one side of the tongue released from the lateral portion of the alveolar ridge (see Doke, 1926; Ziervogel, Louw, & Taljaard, 1976). Thus, the place of articulation for the apical click is only roughly similar to that for /t/ in English and actually more like the Hindi dental stop /t̪/. Nothing even roughly equivalent to the palatal or lateral places occurs in any English stop. The asymmetrical release of the lateral click is, in fact, a very uncommon feature in the world's languages.

We restricted our tests to the voicing and place contrasts among the nonnasalized clicks, for which there are 18 minimal-pair contrasts of either place or voicing. Because we predicted good discrimination, we tried to minimize procedural biases toward good performance. Therefore, we used an AXB discrimination procedure (see *Procedures* section for description) with relatively long interstimulus intervals (ISIs) of 1,000 ms, rather than one with lower memory demands such as 2IAX or 4IAX (see Carney et al., 1977; Pisoni & Lazarus, 1974; Pisoni & Tash, 1974) or with short ISIs of 250 ms or less (Pisoni, 1973; Werker & Logan, 1985). The task was designed so that the matching items were different tokens of a click category rather than physically identical ones (see also Werker & Logan, 1985). Such a task should tap some degree of perceptual constancy for items within a phonetic category. The click syllables were matched across categories for general acoustic properties (e.g., pitch, loudness, and duration) to minimize discrimination on the basis of phonemically irrelevant information. Moreover, the subjects were not given training on the clicks or feedback on the practice trials, as has been done in other studies reporting above-chance

---

[3] The click at the apical place of articulation is similar to the "tsk" sound of disapproval. That at the lateral place is similar to the clicking sound that people sometimes make from one side of their tongue, along with a wink, when flirting or when showing approval or affectionate greeting. It may also occur with a wince, as a sign of regret or frustration, or when urging a horse along. The palatal click does not have a common nonverbal meaning in American society, as do the clicks at the other two places. It is similar, but not identical, to the "tongue cluck," a repetitive vocal play sound of some infants during the second half-year and into the preschool years, and the sound made to represent the clip-clop of a horse's slow gait.

[4] It should be noted that the "auditory experience" argument as presented by Aslin and Pisoni (1980) mentions experience only within a speech context. Of course, the underlying psychoacoustic assumptions of their view could easily be extended to predict that experience with clicks as nonspeech sounds should maintain sensitivity to them even in a speech context. However, if one were to predict systematic effects of nonspeech auditory experience upon developmental changes in speech perception, it would be difficult to decide which sorts of experience would be expected to have an effect and which nonnative contrasts would be affected in what particular directions. For example, would the sounds that the infant makes in early focal development provide such auditory experience? Infant babbling includes not only clicks but also other non-English sounds such as trills and pharyngeal and uvular noises (Stark, 1980) that appear in contrasts that English-speaking adults and older infants find difficult to discriminate. See also *Discussion* section of Experiment 1.

nonnative speech discrimination (e.g., Pisoni et al., 1982; Tees & Werker, 1984; Werker & Tees, 1984b).

Method

Subjects. Nine college students were tested (7 female, 2 male; age range = 19–23 years). All were monolingual American English speakers with no previous exposure to Zulu or other click languages. None had any known hearing or language difficulties. Each was paid $8.00 for participation in a 1½-hr test session.

Stimuli. The test stimuli were selected from naturally produced Zulu click +/a/ syllables recorded by Subject TM, a native Zulu-speaking woman born and raised just south of the Mahlabathini section in the heart of Zululand, South Africa. The accent of people from this region is considered by Zulus to contain the purest pronunciations of the clicks. Subject TM read from a randomized list containing 20 repetitions of each of the 15 clicks. All syllables were produced with high tone. To ensure the desired tonality, examples of bisyllabic imperative verbs with click +/a/ in word-initial position were given for each item on the sequence listing (see Table 1); only the first syllable of the words was spoken. Subject TM was instructed to keep her productions as constant as possible throughout the sequence with respect to duration, loudness, and pitch contour. The utterances were recorded with a Sony T5D portable cassette tape deck, using a directional Audio-Technica microphone.

Table 1
Zulu Words Used for Recordings of Click + /a/ Syllables and Their English Glosses

Click syllable[a]	Zulu verb[b,c]	English gloss[d]
/ʇa/	⟨*ca*ca⟩	be clear
/ʇʰa/	⟨*cha*ya⟩	spread out (v.)
/gʇa/	⟨*gca*ba⟩	make an incision
/ŋʇa/	⟨*nca*ma⟩	give up
/ŋʇa/	⟨*ngca*ma⟩	feast (v.)
/ʗa/	⟨*qa*la⟩	start (v.)
/ʗʰa/	⟨*qha*ba⟩	snap fingers
/gʗa/	⟨*gqa*ba⟩	paint (face) (v.)
/ŋʗa/	⟨*nqa*ba⟩	refuse (v.)
/ŋʗa/	⟨*ngqa*ngqa⟩	shake
/ʖa/	⟨*xa*xa⟩	beat (v.)
/ʖʰa/	⟨-*xha*la⟩	anxiety (n.)
/gʖa/	⟨*gxa*tha⟩	stride (v.)
/ŋʖa/	⟨*nxa*nxa⟩	coax, urge
/ŋʖa/	⟨*ngxa*ma⟩	be angry

[a] Represented in phonetic symbols (see Catford & Ladefoged, 1968; Ladefoged, 1975).

[b] The words are written in current Zulu orthography (which is based on the Roman alphabet), in which ⟨c⟩ corresponds to the apical place of articulation, ⟨q⟩ corresponds to the palatal place, and ⟨x⟩ corresponds to the lateral. An ⟨h⟩ following a click symbol indicates that it is voiceless aspirated, whereas letters preceding a click symbol indicate voicing (⟨g⟩), nasalized voicing (⟨ng⟩), or nasalized voicelessness (⟨n⟩). The voiceless unaspirated items are represented by the click place symbols alone. Thus, for example, the apical click syllables are written as follows: ⟨ca⟩ (voiceless unaspirated), ⟨cha⟩ (voiceless aspirated), ⟨gca⟩ (voiced), ⟨nca⟩ (nasalized voiceless), and ⟨ngca⟩ (nasalized voiced).

Although the nasalized click syllables were recorded, they were not used in the present perceptual experiments.

[c] The speaker produced each of the italicized syllables with high tone, as it is normally spoken in word-initial position in these words (except ⟨-xhala⟩, a suffix in which the syllable of interest is in initial position), which were provided as examples on the sequence list. The initial syllable in all items on the list (including -xhala) is produced with high tone.

[d] All items are imperative verbs, except ⟨-xhala⟩. No bisyllabic imperative verbs beginning with ⟨xha-⟩ exist in Zulu; in fact, we were unable to find any bisyllabic words beginning with ⟨xha-⟩ using high tone.

The nonnasalized click syllables were digitized and stored on disk, using the PCM (pulse code modulation) system of the VAX 11-780 computer at Haskins Laboratories. Author NMS (a native Zulu speaker) eliminated any tokens that were pronounced incorrectly or unclearly or with an incorrect tone or vowel quality. From the remaining syllables, six exemplars of each category were selected for their similarity in length, loudness, pitch, and vowel quality. Preliminary acoustic analyses verified that the selected syllables were physically similar (except for click properties—see second paragraph below). Although there was some degree of variation among the tokens in acoustic properties of the vocalic (vowel) portions of the syllables (e.g., F_0, contour, and amplitude), it was found within as well as between categories, and there was much overlap in these vocalic acoustic properties between categories. The original duration of the selected syllables ranged from 232 to 310 ms, with a mean of 285 ms. The durations of the syllables were modified, by means of a software waveform editor on the Haskins VAX 11-780, so as to restrict the final range to 272–302 ms (M = 288 ms). This was accomplished by iterating or deleting individual pitch pulses from the steady portions of the vowels, or by adding or deleting small amounts of silence in the closure portion of voiceless items. In these cases, Author NMS verified that the editing had not distorted the phonetic properties of the syllables.

F_0 and formant frequency characteristics of the syllables were calculated by linear predictive coding (LPC) analysis using the Interactive Laboratory Systems software (ILS: Signal Technology, Inc., 1986), while voice onset time (VOT) and durations of click bursts were measured by hand marking and measuring the waveform in the waveform editor program. The results are shown in Table 2. The average F_0 contour across the vocalic portions of the items was nearly flat, the overall mean at vowel onset being 202 Hz and at vowel offset being 197 Hz. However, F_0 contour varied slightly along the voicing dimension, because of differences in onset frequency between voicing categories (offset frequency did not differ noticeably). The voiceless aspirated items had a slightly higher starting frequency (M = 212 Hz) than the voiceless unaspirated (M = 204 Hz) or voiced items (M = 190 Hz); the former were slightly falling, whereas the latter two were slightly rising. The F_0 onset difference between the vocalic portions of the voiceless aspirated and voiced click syllables may be akin to that found between English voiceless and voiced stops (Haggard, Ambler, & Callow, 1970).

Acoustic properties of the clicks differ across places of articulation and voicing categories (see Table 2). Place categories differ in duration of click bursts only slightly. They are somewhat longer for lateral clicks (M = 52 ms) than for apical (M = 44 ms) or palatal clicks (M = 43 ms). Voiceless aspirated click bursts are slightly longer (M = 53 ms) than voiceless unaspirated clicks (M = 46 ms), the latter, in turn, being slightly longer than voiced clicks (M = 40 ms). Click amplitude at the peak of the burst varied systematically, being highest for the palatal (M = 52.7 dB signal/noise ratio) and lateral clicks (M = 51.4 dB) and lowest for the apical clicks (M = 39.7 dB). The spectral distributions of the clicks also differed, with the apical click bursts showing relatively greater energy in the high-frequency range than the other two categories, and the palatals showing relatively greater energy in lower frequencies.

VOT, measured as the time between onset of the burst and onset of periodic voicing (Lisker & Abramson, 1964), was longest for voiceless aspirated clicks and shortest for voiced clicks. It should be noted that although we use the click voicing terminology recommended by phoneticians (Catford & Ladefoged, 1968; Doke, 1926),

Table 2
Acoustic Measurements of the Nonnasalized Zulu Click + /a/ Syllables

	Acoustic measures					
Click categories	F_0 onset[a]	F_0 nucleus[b]	F_0 offset	Click duration[c]	VOT[c]	Click amplitude[d]
Unaspirated voiceless						
Apical (/ʇa/)	203	205	199	43.3	62.9	40.55
Range	185–217	196–213	185–204	35–50	36–92	37.4–43.7
Lateral (/ʖa/)	204	200	198	51.7	70.8	50.73
Range	196–213	189–204	189–208	40–60	40–91	46.5–54.2
Palatal (/ʗa/)	207	205	197	44.2	55.8	52.94
Range	189–222	204–208	192–204	35–50	45–71	48.3–54.8
Voiceless aspirated						
Apical (/ʇʰa/)	219	208	201	50	153.8	40.09
Range	213–222	200–213	196–208	35–60	93–148	38.1–42.7
Lateral (/ʖʰa/)	210	190	195	60	143.2	53.5
Range	182–233	192–208	185–204	55–65	134–150	49.6–56.3
Palatal (/ʗʰa/)	208	206	198	47.5	121.4	52.47
Range	196–222	200–217	189–204	45–50	105–140	50.7–54.9
Voiced						
Apical (/gʇa/)	194	190	200	40	33.1	38.3
Range	189–196	185–196	189–204	35–50	29–39	35.6–40.7
Lateral (/gʖa/)	189	187	196	43.3	34.7	49.96
Range	182–196	182–189	185–200	40–50	37–43	46.5–55.0
Palatal (/gʗa/)	186	188	191	37.5	31.1	52.7
Range	175–196	182–192	189–192	35–45	22–40	48.3–54.8

Note. VOT = voice onset time.
[a] Shown in Hz. Measured at vocalic onset.
[b] Measured at vowel nucleus, approximately 80 ms from syllable offset.
[c] Shown in ms.
[d] Shown in dB gain (signal/noise ratio) for a 10-ms window at amplitude peak of click burst.

the VOT durations do not correspond well with the VOT measurements that have been reported for stop voicing categories carrying the same name. In fact, all three click voicing categories involve a lag between burst onset and voicing onset, which would be termed *voiceless* in phonetic descriptions of stop consonants. The lag, even for so-called voiced clicks, is due to the suction mechanism, which prevents release of the velar occlusion (and hence voicing onset) until after release of the click. The VOT durations of voiced clicks do not correspond well with stop voicing categories in English (Lisker & Abramson, 1964). Voiced click VOTs are longer (M = +31–+35 ms) than those associated with voiced stops (M = −102–+21 ms) but shorter than those found in English voiceless stops (M = +58–+80 ms). The voiceless unaspirated clicks have VOTs (M = +56–+71 ms) corresponding to English voiceless stops (but the latter are usually aspirated in initial position). The voiceless aspirated clicks have VOT values (M = +121–+154 ms) far longer than those associated with English voiceless stops. Thus, none of the click voicing categories corresponds well to the acoustic and phonetic properties associated with English stop consonant voicing contrasts.

Procedure. Subjects completed an AXB discrimination test including comparisons for each of the 18 minimal-pair contrasts (see Table 3). Testing was conducted in a sound-attenuated room, with stimuli presented at a comfortable listening level (approximately 75 dB SPL) over Sennheiser HD230 headsets to groups of 2–4 subjects. The stimuli were played out from an Otari MX5050 BQ-II tape deck.

The AXB discrimination test contained 36 blocks of 12 trials each, randomized within blocks. Three stimuli were presented on each trial, and the subjects indicated on a check-off sheet whether the middle stimulus (X) was from the same category as the first (A) or the third (B) stimulus. On the same sheet, they circled a number from 1 to 4 for each trial to indicate their confidence in their answer (1 = *simply guessing* and 4 = *very sure*). The ISIs within trials were 1,000 ms. The intertrial intervals (ITIs) were 6 s, and the interblock intervals (IBIs) were 10 s. Each test block was restricted to a single contrast; there were two test blocks for each of the 18 contrasts, one in the first half of the test and one in the second half. Subjects were given a 10-min break between the first and second half of the test.

On each trial, the middle item was a nonidentical token from the same category as either the first item (A) or the third item (B). We refer to this procedure as name-identity AXB discrimination (see also Werker & Logan, 1985). Subjects were told that on each trial, the first and third items were always from different speech sound categories, even if they didn't sound so, and that the middle item was from the same category as the first or third. They were given 12 practice trials without feedback, which ranged across the 18 contrasts.

After the end of the test, the subjects completed posttest questionnaires, asking them to describe the properties of the syllables they had used to base their discriminations upon and to report how easy they had found the task.

Results

The subjects' confidence ratings were relatively high: A rating of 4 (*very sure*) was indicated for 37% of the trials, and a rating of 3 (*sure*) for 36%. The average rating across the test was 3.04.

The data for percentage of correct performance were entered into a within-subjects analysis of variance (ANOVA), in which the second and third factors were nested within the first factor. The design was 2 (contrast type: place contrast vs.

Table 3
Minimal-Pair Contrasts Among the Nonnasalized Zulu Click Syllables

Feature contrast	Contrast type		
	Place of articulation		
	Apical vs. palatal	Apical vs. lateral	Palatal vs. lateral
Voicing category			
Voiceless unaspirated	/ʇa/–/ʗa/	/ʇa/–/ʖa/	/ʗa/–/ʖa/
Voiceless aspirated	/ʇʰa/–/ʗʰa/	/ʇʰa/–/ʖʰa/	/ʗʰa/–/ʖʰa/
Voiced	/gʇa/–/gʗa/	/gʇa/–/gʖa/	/gʗa/–/gʖa/
	Voicing		
	Voiceless aspirated vs. unaspirated	Voiced vs. voiceless aspirated	Voiced vs. voiceless unaspirated
Place category			
Apical	/ʇa/–/ʇʰa/	/gʇa/–/ʇʰa/	/gʇa/–/ʇa/
Palatal	/ʗa/–/ʗʰa/	/gʗa/–/ʗʰa/	/gʗa/–/ʗa/
Lateral	/ʖa/–/ʖʰa/	/gʖa/–/ʖʰa/	/gʖa/–/ʖa/

voicing contrast) × 3 (feature category: voiceless unaspirated, voiceless aspirated, and voiced for place contrasts; apical, lateral, and palatal for voicing contrasts) × 3 (minimal contrast: voiceless unaspirated vs. voiced, voiceless aspirated vs. voiced, and voiceless aspirated vs. unaspirated for voicing contrasts; apical vs. lateral, apical vs. palatal, and lateral vs. palatal for place contrasts). Table 4 illustrates the design and displays the percentage of correct discriminations on each of the 18 click contrasts. Table 5 lists the significant effects of the ANOVA. The main effect for minimal contrast (within contrast type) indicates that performance was higher on the voiced versus voiceless aspirated distinction than on the other two voicing distinctions and that performance was higher on the place of articulation contrasts that included the palatal click than on those that did not (apical vs. lateral). The Feature Category × Minimal Contrast (contrast type) interaction revealed that regardless of voicing category, performance was lower for the apical versus lateral contrast than for the contrasts involving the palatal clicks and that the voiced versus voiceless aspirated distinction was easiest regardless of place of articulation. Although performance on the other two voicing distinctions was somewhat lower, it did not differ at either the palatal or the lateral places. However, the voiceless aspirated versus voiceless unaspirated distinction at the apical place was the most difficult voicing distinction overall. The main effects of contrast type and of feature category (contrast type) were nonsignificant. There was no generally greater ease with voicing contrasts than with place contrasts.

Discussion

The results indicate that neither a lack of experience hearing clicks in spoken English nor their phonological irrelevance had a negative effect on click discrimination. They are consistent with our prediction that discrimination should be easy for nonnative contrasts that cannot be assimilated to English phoneme categories.

Experience listening to clicks as nonspeech (see footnotes 2 and 3) might also explain the maintenance of perceptual sensitivity, however. This possibility is weakened by observations of other nonnative contrasts for which highly similar nonspeech experience should be relevant. For example, the intervocalic Spanish trill versus flap contrast /ɾ/–/r/ ([r]) is assimilated to English /t/ or /d/, or to /r/ ([ɹ]) by English-speaking adults, and is apparently lost by the second half-year of life in English-learning infants (Eilers et al., 1982; Oller & Eilers, 1983). Yet [r] is similar to the rolling tongue trill that is used in infants' vocal play, in children's imitations of airplane or car sounds, and in mimicry of the cat's purr, while [ɾ] is a common allophonic variant of American English /t/ or /d/ (intervocalically) or of British /r/.

Table 4
Mean Percentage of Correct Performance on the Minimal-Pair Click Contrasts

Feature contrast	Contrast type		
	Place of articulation		
	Apical vs. palatal	Apical vs. lateral	Palatal vs. lateral
Voicing category			
Voiceless unaspirated	97.7	80.6	95.8
Voiceless aspirated	97.2	82.9	94.4
Voiced	92.6	86.6	96.8
	Voicing		
	Voiceless aspirated vs. unaspirated	Voiced vs. voiceless aspirated	Voiced vs. voiceless unaspirated
Place category			
Apical	82.4	99.1	88.0
Palatal	88.0	98.2	89.4
Lateral	89.4	97.7	89.4

Table 5
Significant Analysis of Variance Effects for Experiment 1

Variables	*df*	*F*	*p*
Minimal contrast (within contrast type)	4, 32	17.95	< .0000
Feature Category × Minimal Contrast (within contrast type)	8, 64	2.40	< .025

Another potential explanation might be that the clicks are psychoacoustically robust and thus resistant to decline in discriminability (Burnham, 1986). Although the very fact that the clicks are easily discriminated fits Burnham's definition of psychoacoustic robustness, one would prefer to use independent criteria. For example, Burnham hypothesized a correlation between the robustness of a contrast and its representation in world languages. Although the click contrasts are relatively rare, this is not a serious problem for the psychoacoustic argument. The linguistic distribution of a contrast would presumably also be influenced by other factors, such as articulatory ease or sociocultural forces. Robustness may be necessary for widespread adoption of a contrast, but it is not sufficient.

The click distinctions might nonetheless satisfy other criteria for robustness, which might explain the variations in discrimination among the click voicing distinctions and among the place of articulation distinctions. Burnham (1986) argued that at least some non-English *stop* voicing distinctions are psychoacoustically robust (e.g., prevoiced [b] vs. voiceless unaspirated [p]). Indeed, nonnative stop voicing contrasts are relatively amenable to training (e.g., Aslin & Pisoni, 1980; Werker & Tees, 1984b). The voiced versus voiceless aspirated click contrast, which yielded the highest discrimination performance in the present results, would certainly seem to be psychoacoustically more robust than the others, given that it involves the largest VOT separation. Such a psychoacoustic explanation would be compatible with our hypothesis that nonassimilable contrasts should be discriminated on the basis of their auditory (or phonetic) properties.

Compared with voicing, nonnative place of articulation contrasts have generally proved more difficult perceptually and resistant to training (e.g., Tees & Werker, 1984; Werker & Tees, 1984b). Yet the click place contrasts were also quite easy to discriminate. Allophonic considerations cannot account for the pattern of click place discriminations. The presence of a phonetic feature in the apical click that is roughly similar to one found in English /t/ did not provide special perceptual aid. Instead, discrimination performance was best for place distinctions involving the palatal click, which differs in place of articulation from any English phoneme. Thus, the acoustic properties of the stimuli played a larger role in place discriminations than did their phonetic properties. The amplitude variation across the three click places is the most obvious acoustic difference. Specifically, the palatal place is associated with the highest amplitude click burst. The apical place, although most phonetically similar to English, was associated with the lowest performance and the lowest amplitude click burst.

The psychoacoustic approach thus appears to handle discrimination performance on click distinctions. However, it should also be able to explain why performance on these contrasts reached higher levels than it has on other contrasts identified by Burnham. Although the clicks differ across places of articulation in their spectral distributions and amplitudes, and slightly in their durations, variations in burst amplitude and spectrum also differentiate the Hindi retroflex versus dental stop (/ṭ/–/t̪/) contrast (Tees & Werker, 1984; Werker et al., 1981; Werker & Logan, 1985; Werker & Tees, 1984a, 1984b). Those authors found the retroflex–dental stop discrimination to be very difficult for English-speaking adults even after training, although low memory-demand conditions (unlike the conditions of the present study) did lead to improved performance. Similarly, although the bursts of the Thompson /k'/ versus /q'/ contrast appear to differ somewhat in amplitude and duration, that place contrast is also difficult for English-speaking adults, even with instructions or training (e.g., Werker & Tees, 1984b). Listeners indicate hearing both members of the Hindi /ṭ/–/t̪/ contrast as /d/ while hearing both members of the Thompson contrast as /k/ (Werker & Logan, 1985). It is this phonemic influence for the Hindi and Thompson contrasts, we argue, that is responsible for their poor discriminability. We suggest that the high performance on the Zulu clicks occurred because no such phonemic influence appears to have operated for them, thus permitting subjects more direct perceptual access to their auditory (nonspeech) or phonetic (articulatory) properties. The implication that phoneme perception may supercede perception of purely auditory or phonetic information in the signal is not new. It is supported by research on categorical perception of speech sounds in general (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), as well as by perceptual constancy (e.g., Kuhl, 1980) and phonetic trading relations in perception of phoneme categories (e.g., Best, Morrongiello, & Robson, 1981) and by recent demonstrations that speech perception takes precedence over auditory perception of the same signal (e.g., Whalen & Liberman, 1987).

Subjects' answers on the posttest questionnaires indicated a virtual failure to assimilate the clicks to English phonemes. All subjects stated that they relied on auditory (nonspeech) properties of the sounds (e.g., "clicks," "plops," "pops," "percussion instruments," "water drip," "finger snap," and "clap") when discriminating the syllables. Interestingly, several subjects also indicated relying on articulatory (phonetic) differences (e.g., "tongue popping," "tongue clucking," and "sounds coming from different areas of the mouth"). Although some also thought there might be secondary vowel quality, intonational, or loudness differences between the syllables, they indicated that these were small and difficult to differentiate. Only 2 subjects related any of the sounds to English consonants, and both indicated that they were able to use these consonantal associations for only a couple of test blocks. One referred to <d–t> differences (most likely associated with the apical voiced vs. voiceless aspirated distinction), and both referred to <k–g> differences (most likely associated with the lateral voiced vs. voiceless aspirated distinction).

We believe that the very high levels of performance in Experiment 1 reflect a perceptual focus on the auditory and/

or phonetic properties of the clicks. However, this may not occur solely because their acoustic differences are psychoacoustically robust but rather because a failure to assimilate the clicks to English phonemes results in a perceptual focus on their nonphonemic properties. The amplitude differences among the click place contrasts would seem the most likely source of a robust psychoacoustic difference. Therefore, in Experiment 2, we tested whether adults would still discriminate a click place distinction after the click amplitudes were equated, on the basis of the remaining acoustic differences.

Experiment 2

In Experiment 2 we compared American listeners with Zulu listeners. This allowed us to determine both whether amplitude modifications of the clicks had distorted crucial phonetic properties of the syllables according to native listeners and whether differential linguistic experience influenced discrimination on the basis of the remaining acoustic properties. To provide the best chance of observing a developmental reorganization in click discrimination by infants (see Experiment 3), as would be predicted by Werker's findings (Werker et al., 1981; Werker & Tees, 1984a), we chose the place contrast on which adults had shown the lowest performance in Experiment 1, the voiceless unaspirated apical versus lateral distinction. This click contrast is represented in phonetic symbols as [ʇ] versus [ʖ], and the syllables are written in the Zulu orthography as <ca< versus <xa>. In the original stimuli, the /ʖ/ click burst was higher in amplitude than /ʇ/ on oscillographic tracings (see row a of Figure 1) and in the acoustic analyses (Table 2).

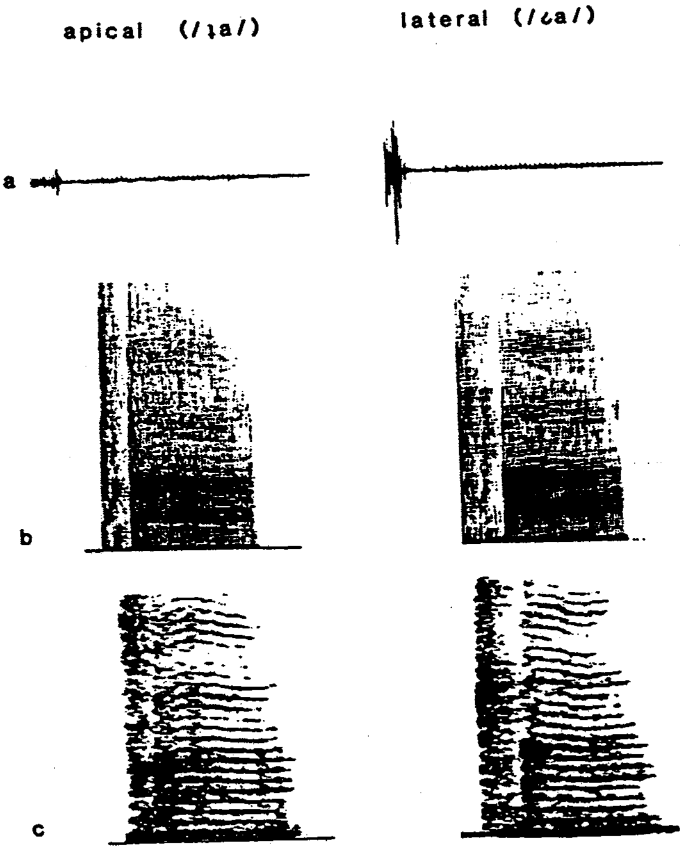

Figure 1. Panel a: oscillographic displays of an original (amplitude unmodified) voiceless unaspirated apical click (/ʇa/) and lateral click syllable (/ʇa/). Panel b: wide-band spectrograms of the amplitude-equated versions of those syllables. Panel c: narrow-band spectrograms of the amplitude-equated versions of those syllables (see Experiment 2).

Method

Subjects. Eight (4 male, 4 female) monolingual English-speaking college students (age range = 19–22 years) formed the English language group.[5] Six additional students (2 males, 4 females) formed the Zulu language group (age range = 19–36 years). All of the latter group had been born and raised in South Africa but were currently enrolled in colleges in New England. Author NMS was one of the Zulu subjects. All in the Zulu group spoke English fluently. Three were from Zulu-speaking areas of South Africa and had learned Zulu as their first language. The other three were from Xhosa-speaking areas and had learned Xhosa as their first language, but also spoke Zulu fluently. It should be noted that Zulu and Xhosa are very closely related, both being Bantu languages spoken by the Nguni peoples of South Africa. Speakers of one language can generally understand conversation in the other, although many vocabulary items are unique to one or the other language. The click system is identical to that of Zulu.

None of the subjects had any known hearing or language difficulties. Each received $4.00 for 30–45 min of participation.

Stimuli. The amplitudes of the click bursts were equated across the two categories (in terms of dB gain levels at the peak) by means of software waveform editing, by reducing the amplitude of the /ʖ/ click bursts (but not the vocalic portion) and increasing that of the /ʇ/ clicks. The perceived loudness of the clicks was constant across the two categories. According to Author NMS, the modified /ʇ/ clicks sounded "wet" and the modified /ʖ/clicks sounded somewhat attenuated or "swallowed," but the changes did not interfere seriously with their phonemic category membership.

The remaining acoustic properties of the syllables are listed in Table 6. The between-category distinction appears to be marked primarily by differences in F_1 and perhaps F_2 transitions (see Figure 1, rows b and c) and in spectral distribution of the clicks. The rising F_1 transitions are more rapid for /ʖ/ than /ʇ/, although the magnitude of the frequency excursion is similar. Frequency of F_1 at the beginning and asymptote of the transition is higher for /ʖ/. Both categories show a slightly rising F_2 transition, but the onset frequency is higher for /ʇ/. For both categories F_3 is nearly flat.

On the basis of discrete fourier transform (DFT) analyses of the click bursts, which present Intensity × Frequency information about brief portions of the signal, the tilt of the power spectrum differs at the highest amplitude portions of the clicks (see Figure 2). The /ʇ/ (apical) clicks show a rising energy distribution with a secondary concentration of low frequency energy, whereas /ʖ/ clicks show both high and mid-range frequency peaks. There are also between-category differences in frequency characteristics of the onset and offset transients of the clicks. The onset transients are biased toward higher frequencies in /ʇ/. The offset transients for /ʇ/ show two distinct concentrations somewhat below the F_1 onset frequency and near the F_2 steady-state, while /ʖ/ shows offset transients with energy concentrations near F_3 and F_4.

---

[5] These also participated in Experiment 3, as a comparison group for infants.

Table 6
Acoustic Measurements of the Amplitude-Modified Voiceless Unaspirated Apical and Lateral Click Syllables

Measure	Apical (/ʇa/)		Lateral (/ʖa/)	
	Hz	Range	Hz	Range
Acoustic measures				
Vocalic portions				
F_1 onset	699	627–824	795	507–995
F_1 transition asymptote	1147	1113–1175	1232	1037–1380
F_1 excursion depth	448	329–529	437	282–680
F_2 onset	1557	1420–1666	1427	1022–1571
F_3 steady-state	2742	2601–2796	2772	2642–2816
F_0 onset	203	185–217	204	196–213
F_0 at vowel nucleus[a]	205	196–213	200	189–204
F_0 offset	199	192–204	198	189–208
Clicks				
Frequency peaks				
Low	4655	4586–4736	4355	3971–4661
High	119	109–126	2453	2138–2843
Onset transient peaks				
Low	4557	4236–4810	4529	4066–4661
High	3248	1443–4023	2485	2115–2739
Offset transient peaks				
Low	1538	1322–1609	4476	4275–4741
High	440	431–454	2420	2190–2509
Durational measures (in ms)				
Syllable length	286	274–296	293	286–302
Click duration	43	36–54	52	39–64
VOT	63	36–91	71	40–91
F_1 transition duration	64	50–80	30	20–50

Note. VOT = voice onset time.
[a] Measured at approximately 80 ms from vocalic offset.

The duration of the syllables and of the clicks alone differed only very slightly and showed much overlap between categories. Closure durations were nearly identical, as were F_0 contour and level.

Procedure. The English language group was tested under the same experimental setup as described for Experiment 1. The members of the Zulu group were tested, three at a time, in a quiet room near their college. The tests were presented to the latter group over the built-in speaker of a portable Sony T5D cassette tape deck at a comfortable listening level (approximately 75 dB SPL). All listeners completed the test(s) in a single session.

Both language groups completed a name–identity AXB discrimination test, of the same format as described for Experiment 1, except that it contained only trials with /ʇ/–/ʖ/ tokens and consisted of six blocks of 12 trials. For the English language group only, the first block of 12 trials served as a no-feedback practice set; their answers on this block were not scored. The Zulu language group first completed a standard identification test on the modified stimuli. This consisted of five blocked randomizations of the 12 stimuli, presented singly, with 2.5-s ISIs and 4-s IBIs.

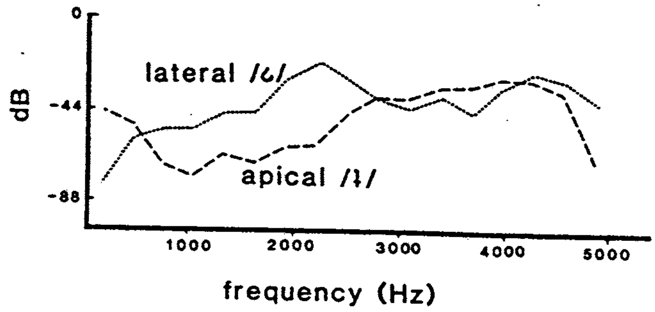

Figure 2. Spectral sections of the highest-amplitude portions of the click bursts from an amplitude-modified voiceless unaspirated apical (solid line) and lateral (dotted line) click syllable (see Experiment 2).

Results

Identification task, Zulu language group. Overall, the Zulu listeners found the modified syllables to be acceptable tokens of apical and lateral voiceless unaspirated clicks. The group labeled the tokens correctly on 92% of the trials. Author NMS's performance was not noticeably different from that of the other Zulu listeners. Performance was somewhat higher for /ʖ/ tokens (M = 98% correct) than for /ʇ/ tokens (86%), suggesting that the latter may have been somewhat less acceptable phonetically. However, this difference did not reach standard levels of significance according to ANOVA ($p = .14$). The Xhosa speakers had more difficulty with the /ʇ/ items (M = 79% correct) but only very slightly more difficulty with the /ʖ/ items (M = 96%), than did the Zulu speakers (M = 92% and 100%, respectively). Post hoc t tests of differences between the language subgroups on both comparisons failed to reach significance ($ps > .20$). However, the small ns restrict confidence in the null results of these statistical tests.

AXB discrimination, both language groups. The data for percent of correct answers on the AXB name–identity discrimination tasks were entered into a one-way ANOVA with language group as the between-subjects factor. For this purpose, the Zulu and Xhosa listeners were collapsed into a single Zulu language group. Although mean performance was somewhat higher for the Zulu language group (87% correct: 84% for Xhosas and 89% for Zulus) than for the English language group (78% correct), the difference was not statistically significant ($p = .12$). However, this null result is qualified, again, by the small n of the Zulu group, as well as by language group differences in listening conditions and in initial practice with the tokens.

Discussion

The results of Experiment 2 indicate that Zulu speakers can identify, and both Zulu and English speakers can easily discriminate, the voiceless unaspirated apical versus lateral click contrast on the basis of acoustic distinctions other than click amplitude differences. Moreover, the English language group in the present study did virtually as well with these amplitude-modified stimuli (78% correct) as the listeners in Experiment 1 had done with the unmodified version of this contrast (80%), suggesting that amplitude variations did not play a large role in perception of the contrast in the earlier study. Further research involving parametric variations of the remaining acoustic cues (e.g., via digital resynthesis) in amplitude-equated click syllables could determine their relative importance in click perception by native and nonnative listeners.

The possibility for developmental reorganization in perception of Zulu clicks by English listeners, however, still remained. Although adults are well able to discriminate click distinctions, infants might show a decrement in discrimination of the amplitude-modified clicks at the same age as they have for other nonnative contrasts (e.g., Werker et al., 1981; Werker & Tees, 1984a). If so, it would indicate that different processes underlie the responses of adults versus infants older than 10–12 months of age to nonnative speech contrasts. For example, whereas adults' perception shows clear phonemic influences (i.e., assimilating the nonnative sounds to native phonemes where possible and otherwise focusing on auditory or articulatory information), infants' perception at 10–12 months may show simpler language-identity influences (i.e., maintaining attention only to sounds that are familiar in the language environment). In this scenario, infants would simply tune out or stop attending to differences between the unfamiliar, strikingly non-English, clicks by around 10–12 months. On the other hand, infants might perceive the clicks similarly to the adults and thus continue to discriminate click contrasts even at 10–12 months and beyond.

Experiment 3

To test for these possibilities, we conducted a cross-sectional study of /ʇa/–/ʖa/ discrimination by 6–8- month-, 8–10-month-, and 10–12-month-old infants for comparison with Werker's findings on other nonnative contrasts. To increase our opportunity to observe some perceptual decline for the click contrast, we added a fourth group of infants from 12 to 14 months old. We also tested discrimination for the native English stop place contrast, /ba/–/da/, as a control comparison. For this study, we adapted an operant-conditioning visual habituation procedure (Miller, 1983) (see *Procedure* section for description). It should be noted that this procedure differs from the conditioned head-turn paradigm used by Werker and colleagues (Werker et al., 1981; Werker & Tees, 1984a). We chose to use the visual fixation technique because it appears amenable to testing infants across a wider age range (at least from 2–13 months) than either the nonnutritive sucking technique (useable only between 1 and 4 months; e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971) or the visually reinforced head-turn technique (trainable only after 5–6 months; e.g., Eilers et al., 1979; Werker et al., 1981). We hoped it would be thus useful for future developmental studies.

Method

Subjects. The subjects were 40 infants from middle and upper middle-class homes. A total of 62 infants was tested. Twenty-two infants were eliminated from the final data set for the following reasons: crying (8), inattention to the visual fixation target (6), equipment failure (4), experimenter error (1), inability of the observer to determine accurately the infants' fixations of the visual target (2), and performance levels that were more than 2 standard deviations beyond the means for the infant's age group (1). This left 10 infants each in the following age groups: 6–8 months (27–35 weeks), 8–10 months (36–44 weeks), 10–12 months (45–52 weeks), and 12–14 months (53–61 weeks). The mean age for each group was, respectively, 29.3 weeks, 39.2 weeks, 48.3 weeks, and 59.7 weeks. There were approximately equal numbers of boys and girls at each age. All were being raised in monolingual English environments. All had normal deliveries following normal full-term pregnancies. All infants were in good health at the time of testing; none was on medication. None had any known hearing problems.

For comparison purposes, the 8 American adult subjects from Experiment 2 were also tested on the English and Zulu contrasts in the visual habituation procedure (see *Procedure*).

Stimuli. The Zulu discrimination test of the present experiment used the six amplitude-modified tokens each of /ʇa/ and /ʖa/ used in Experiment 2. The English discrimination test used six tokens each of /ba/ and /da/ produced by an adult male speaker of American English. The latter stimulus set was developed by Werker and colleagues (e.g., Werker et al., 1981) as their native-language control contrast, for which infants show good discrimination at all ages tested, as expected. The multiple tokens for each category thus provided a test of some degree of within-category perceptual constancy.

The stimuli for each test were recorded on a separate tape in such a way that a randomized sequence of the six items in one category appeared on one channel of the tape while the items in the opposing category appeared on the other channel, with exactly synchronous onsets. This permitted smooth switching between stimulus channels at the stimulus shift point during the test (see *Procedure*). The interstimulus intervals were 750 ms. Each tape contained approximately 45 min of continuously recorded stimulus presentations.

Procedure and Apparatus. All subjects were tested on both the Zulu contrast and the English contrast, during a single session. The infant-controlled visual habituation technique for assessing auditory discrimination was developed by Horowitz (1975) and adapted by Miller (1983) to test for perceptual constancy within categories. Generally speaking, in this paradigm an initial series of auditory stimuli (the "familiarization" set) is presented contingent on the subject's fixations of a projected visual pattern. Habituation to the auditory familiarization stimuli is indexed by a decrement in visual fixation to a criterion level, at which point a shift to a new set of auditory stimuli (the "test" set) occurs. A significant postshift recovery of fixation time to the same visual target indicates discrimination between the auditory familiarization and test stimuli.

Subjects were tested in a sound-attenuated room adjacent (and connected by one-way windows) to the control room from which the observer and experimenter conducted the session. The visual fixation slide was back-projected (15 cm × 15 cm) through a translucent rectangular sheet of acetate slightly larger than the projected image, which was affixed in the center of a one-way window. The remainder of the window was covered with opaque black material except for a small peephole (invisible to the subject) by the lower right corner of the projected image. Two slides were used, one for each of the language tests. Each showed a 4 × 4 checkerboard, one of blue and white and the other of yellow and green (equated for brightness), in the center of which was the broken outline of a circle in orange or red, respectively.

A small booth in the testing room was attached to the top and sides of the projection window. The booth was covered inside with black felt, including the ceiling. The walls of the booth measured approximately 1¾ m high × 1 m wide; the opening was approximately 1 m wide. Each adult or infant subject was positioned in the booth, with eyes approximately 45 cm from the projected slide. Adult subjects were seated on a chair, infants sat either in an infant seat stabilized on a small table (younger infants) or in a highchair (older infants) inside the booth. Parents of the infant subjects sat quietly behind or to the side of the booth out of the infant's view, except when the infant would not tolerate the seat or chair. In the latter cases, the parent held the infant in a sitting or standing position within the booth, so that the infant could not see the parent's face. Parents were cautioned not to talk to or in any way distract their infants or bias their responses during the test sessions. During testing, the parents wore Sennheiser HD230 closed-model headsets through which music was played to prevent them from hearing the audio stimuli.

An observer, who also listened to music over headsets, viewed the subject's fixations of the projected checkerboard (as judged by corneal reflection) through the peephole. The observer depressed a "looking" key whenever the subject fixated the checkerboard; there were also buttons to depress whenever an infant subject was crying or sleeping. This information was all recorded by an Atari 800 computer, which was programmed to end each visual fixation trial by closing a Gerbrands shutter on the slide projector whenever the subject looked away from the slide for more than 2 s. The Atari reopened the shutter after a 1-s ITI, and a new trial began. The computer terminal displayed commands to an experimenter, who could not see the subject. These were commands to play the audio stimuli whenever the subject fixated the checkerboard, to stop audio presentations whenever the subject looked away from the slide, and to switch from the "familiarization" stimulus channel to the "test" channel when the computed habituation criterion was reached. The experimenter stopped the tape, and switched channels, only during the silent ISIs between syllables. The habituation criterion was based on Miller's (1983) formula: a fixation-time decrement of 50% or more on two consecutive trials, relative to the average of the two longest-duration trials for the familiarization phase of that test. The program calculated and updated the habituation criterion on every trial. After habituation had been reached and the stimulus shift had been signaled, the session then continued with "test" stimulus presentations, until the habituation criterion was met again. The computer automatically terminated the session at this point or whenever any infant accumulated 30 s of crying or sleeping during a session.

The auditory stimuli were played to the subjects from the Otari tape deck, through a Kenwood amplifier and a specially constructed listening station that permitted easy switching of audio channels, and into a Jamo compact loudspeaker centered over the projection panel, above the ceiling of the booth. The stimuli were played at approximately 75 dB SPL.

Upon arrival at the laboratory, infant subjects were given a 5–10-min period of acclimation while the procedures were briefly explained to the parent. Parents completed a permission form and a background questionnaire on family language background and on medical characteristics of the pregnancy and delivery. Each infant then participated in the two discrimination tests, one for Zulu (/ɬa/–/ɮa/) and one for English (/ba/–/da/). The order of test presentations was counterbalanced across infants, with approximately equal numbers of infants at each age in each test order. The infants were given a 5–10-min break (or longer, if needed) between tests in order to minimize any carryover of habituation from the first to the second test.

The adults were given brief verbal instructions prior to the task, which indicated that there were both between- and within-category variations among the stimulus sets and that they should listen as long as they wished (dependent on fixating the slide) until they felt familiar with the range of variation. Because the task is rather unusual for adults, all these subjects were tested with English first to familiarize them with the procedure.

Table 7
Significant Analysis of Variance (ANOVA) effects for Experiment 3

Effect	*df*	*F*	*p*
ANOVA effects, infants			
Cumulative fixation during familiarization			
Language × Test Order	1, 32	4.50	< .04
Habituation during familiarization			
Habituation	1, 32	53.41	< .0000
Preshift versus postshift means			
Recovery (pre vs. post)	1, 32	37.00	< .0000
Language	1, 32	3.07	< .09
Language × Test Order	1, 32	16.73	< .0003
Recovery × Age × Test Order	3, 32	3.24	< .03
Zulu alone	1, 32	33.55	< .0000
English alone	1, 32	10.21	< .003
Difference scores			
Age × Test Order	3, 32	3.24	< .03
Age × Test Order, Zulu alone	3, 32	3.30	< .03
ANOVA effects, Adults			
Preshift versus postshift means			
Recovery (pre vs. post)	1, 6	24.25	< .003

Results

Infants. Three preliminary ANOVAS were conducted on the familiarization-phase data to assess whether there were any important language-related differences in habituation. Significant effects of all ANOVAS are listed in Table 7. A 4 (age) × 2 (language) × 2 (test order) ANOVA was conducted on the data for total number of trials taken to reach the habituation criterion. The effects were all nonsignificant, indicating no systematic language-related variation on this measure of habituation. Another 4 × 2 × 2 ANOVA was conducted with the data on the cumulative fixation time before reaching the habituation criterion. A significant Language × Test Order interaction indicated a larger cumulative fixation time during the familiarization phase of the first test for English (M = 61.03 s) than for Zulu (M = 46.15 s). Cumulative fixation during familiarization on the second test was lower and did not differ substantially between languages (Ms = 41.64 and 39.19 s, respectively). This pattern indicates that the infants preferred listening to the English syllables more than the Zulu during the first test but that this difference disappeared by the second test, probably because of a general response decrement (though small) across the test session. A third 4 × 2 × 2 × 2 (habituation trials) ANOVA tested for differences in extent of habituation. The habituation trials effect indicated a significant difference between mean fixation for the two familiarization trials with highest fixation durations compared with the mean for the two trials just prior to stimulus shift, that is, significant habituation (see Figure 3).

For the analyses on discrimination performance, the mean fixation time for the last two preshift trials and the mean for the first two postshift trials[6] were computed, following Miller's (1983) approach for analyzing data collected in the visual fixation techniques. A preshift–postshift comparison measures the occurrence and degree of stimulus discrimination as extent of immediate postshift recovery in conditioned fixa-

---

[6] The postshift trials were determined relative to the first time the infant fixated the slide, and thus heard at least one shift stimulus, after the habituation criterion had been reached. This definition is necessary because a stimulus shift has not occurred for the subject unless some audio shift stimuli have actually been presented. Some infants habituated to 0 fixation time and then continued without any fixations during the first several slide presentations after the shift phase had been begun by computer (the shutter opens automatically after 1-s intertrial intervals, then closes automatically after 2 s of slide presentation without fixation, and continues to cycle through this way until the infant looks at the slide). For these infants, the first postshift trial with less than 0 fixation was considered to be the de facto first postshift trial.

tion. It is similar to the data analysis approach for the high-amplitude-sucking habituation technique (e.g., Eimas et al., 1971; Streeter, 1976). These data were first entered into a 4 (age) × 2 (sex) × 2 (language) × 2 (test order) × 2 (recovery: preshift vs. postshift) ANOVA (see means, Table 8 and Figure 3) to determine whether we could collapse the data across sex. The results indicated no systematic effects of sex, so the ANOVA was recomputed without sex as a factor (see Table 7). The main effect of recovery indicated that the postshift fixation times were longer than the preshift times; that is, the change was discriminated. There were no other significant main effects in the overall ANOVA, although the language effect approached significance, suggesting a trend toward higher fixation times to English ($M = 4.94$ s) than to Zulu ($M = 3.78$ s). However, note that this refers to the mean of preshift and postshift fixations. Therefore, it does not indicate language differences in discrimination, but rather may suggest something like a higher general interest level in English, as found for cumulative fixation during familiarization. Indeed, recovery was significant for each language separately (Table 7). There was no significant Language × Recovery interaction, indicating that discrimination was not consistently stronger for English than for Zulu.

There were two other significant interactions in the overall ANOVA. The Language × Test Order interaction showed the same pattern in the averaged preshift and postshift fixation times as had been found in the analysis of cumulative fixation during familiarization: higher values for English when it was tested first ($M = 6.35$ s) than for Zulu ($M = 4.38$ s) and lowest values for both when tested second (*M*s = 3.14 and 3.25 s, respectively). This pattern did not reflect a difference in discrimination performance, though, because the Language × Test Order × Recovery interaction was not significant. It may instead reflect (again) an overall attentional preference for English sounds.

The Recovery × Age × Test Order interaction appeared to suggest that the postshift recovery in the oldest group (12–14 months) was greater for both languages when English rather than Zulu was tested first but that the opposite held true for the three younger groups. In order to simplify the description of this three-way interaction and to verify the interpretation, another 4 (ages) × 2 (languages) × 2 (test orders) ANOVA was run on a measure of postshift recovery magnitude, calculated as difference scores (postshift fixation − preshift fixation). The Age × Test Order interaction was significant, because the difference scores are simple transformations of the data in the overall ANOVA. According to this simplified measure, also, the oldest group appears to discriminate better across both languages when tested on English first, whereas the younger three groups showed better discrimination when tested on Zulu first (see Figure 4). However, this interaction is essentially uninterpretable because Neuman-Keuls tests failed to reveal any significant pairwise differences among the data points and also because the language difference was confounded with sex of speaker. This Age × Test Order interaction was significant only for the Zulu tests.

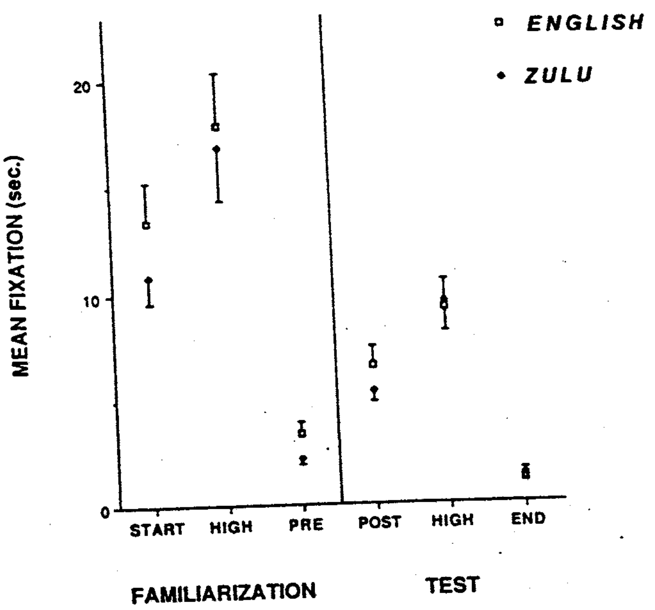

Figure 3. Habituation of infants' visual fixations during the familiarization phase and dishabituation following the stimulus shift (test phase), for tests with the Zulu click and English stop contrasts in Experiment 3. (Each data point represents the mean fixation time, across ages, of a two-trial block. Start = first two trials; high = two trials with highest fixation times, for familiarization or test phase; pre = the two trials immediately preceding the stimulus shift; post = the first two stimulus-shift trials; and end = last two trials of the test phase. The vertical bars extending from each data point represent standard error scores. It should be noted that standard habituation curves averaged across all trials of all infants could not be computed since the infant-controlled visual fixation procedure yields variable numbers of trials for individual subjects. Also note that because of rapid habituation, [as few as four familiarization or four test trials] in some infants, there is partial or total overlap in the test phase for post and high data, or in the familiarization phase for start and high data, for approximately one third of the subjects on the Zulu or the English test.)

Table 8
Mean Fixation Times for Preshift and Postshift Trial Blocks, for Each Age Group on Each Language Test

	English		Zulu	
Age	Preshift	Postshift	Preshift	Postshift
6 months				
M	5.09	6.45	2.82	6.39
SE_m	1.56	1.61	0.57	1.07
8 months				
M	3.09	6.55	1.74	4.99
SE_m	0.87	2.13	6.47	1.23
10 months				
M	3.20	5.01	2.04	4.11
SE_m	0.94	1.04	0.60	0.84
12 months				
M	2.24	7.89	2.21	5.91
SE_m	0.61	2.50	0.67	0.84
Adults				
M	6.50	20.13	9.95	15.45
SE_m	1.51	4.15	1.66	2.29

Note. Preshift values are means for last two trials prior to stimulus shift. Postshift values are means for first two trials of test stimulus presentations.

Adults. To assess the discrimination performance of the adults on the visual habituation task, the mean of the last two

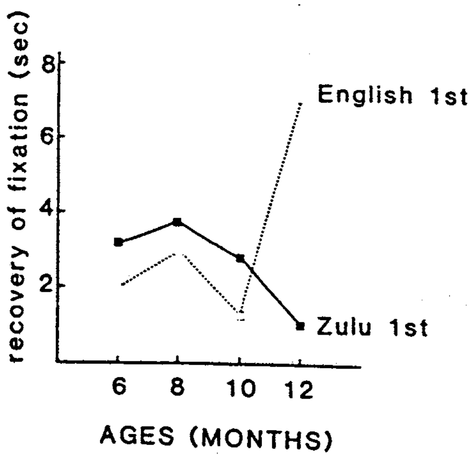

Figure 4. The Age × Test Order interaction found in the difference scores for the infants in Experiment 3. (The ordinate represents the magnitude of fixation-time recovery in response to the stimulus change, computed as the mean fixation time on the first two postshift trials [test phase] minus the mean fixation time during the last two preshift trials [habituation criterion of the familiarization phase].)

familiarization trials (preshift) and the mean of the first two postshift trials were computed, as for the infants. These data were entered into a 2 (sex) × 2 (languages) × 2 (recovery: preshift vs. postshift) ANOVA. The only significant effect was recovery (Table 7), which indicated discrimination of the stimulus change across both languages. The Language × Recovery interaction was only marginal, suggesting a trend toward a greater discrimination of the English than of the Zulu "test" stimuli. Because the adults all received the same test order (English first), test order effects could not be assessed.

Discussion

The results indicate that in the visual habituation paradigm, both English-environment infants and adults discriminate the Zulu click category distinction between the amplitude-modified tokens of /ʇa/ and /ʖa/ as well as they do the English /ba/–/da/ contrast.

There was no significant infant age change in discrimination of the English versus the Zulu click contrasts, as would be predicted by both the auditory experience proposal (e.g., Aslin & Pisoni, 1980), and the specific phonological relevance proposal (Werker et al., 1981; Werker & Tees, 1984a). It should be recalled, in this context, that we used a different procedure than did Werker and colleagues, as well as a different nonnative contrast. Moreover, whereas Werker's headturn procedure uses 1,500-ms ISIs, we used 750-ms ISIs, possibly involving lower memory demands, which could lead to improved performance. The possibility that differences between our infant findings and those of Werker and colleagues were caused by methodological differences needs to be tested. However, we suspect that the ISI differences are not crucial, because 750 ms is in the long ISI range for adults, which typically leads to phonemic-level perception of speech (Werker & Logan, 1985), and in fact other infant researchers have found significant language differences in infant speech perception with even shorter ISIs (400–600 ms) in the headturn procedure (e.g., Burnham, 1986; Eilers et al., 1977, 1979) and the sucking-rate habituation procedure (e.g., Streeter, 1976). Furthermore, the visual fixation technique has been shown to be sensitive to age and stimulus differences in infants' perception of speech qualities (Miller, 1983).

The current results are consistent with the argument that the nature of the developmental change found by Werker and colleagues at 10–12 months for discrimination of (assimilable) nonnative contrasts is a transition toward perceiving speech sounds in relation to native language-specific phonemic contrasts. The similarity in the pattern of the infants' and adults' responses in the current visual habituation paradigm further suggests a commonality in their approaches to the Zulu click contrasts. We suggest that for both age groups this entails a failure to assimilate clicks to native phoneme categories.

General Discussion

The overall pattern of results from the present three experiments is consistent with our prediction that discrimination ability should remain high throughout infancy and in adulthood for nonnative contrasts that are unlikely to be assimilated to any native phonemic categories. This prediction was based on the reasoning that phonemic perception entails assimilation of nonnative speech sounds to native categories whenever possible but that when they are not assimilated, perception focuses either on purely auditory or phonetic (articulatory) properties.

We argued earlier that neither the psychoacoustic hypothesis (e.g., Burnham, 1986) nor the allophonic hypothesis (e.g., Tees & Werker, 1984; Werker et al., 1981) alone could fully account for variation in developmental reorganization for the discrimination of nonnative contrasts. However, each may account for a different portion of the variation. Specifically, Tees and Werker suggested that allophonic experience maintains some degree of perceptual sensitivity for phonetically similar contrasts. Although that argument cannot account for the present findings with Zulu clicks, it may nonetheless apply to variations in performance on single-category and opposing-category contrasts, and possibly category-goodness difference contrasts (see introduction)—respectively, those that are assimilated to a single native phoneme, those that are assimilated to a native contrast, and those for which one member is better assimilated to a native category than is the other. On the other hand, Burnham's psychoacoustic proposal may apply most clearly to variations in nonassimilable contrasts like the clicks. That is, psychoacoustic influences may be most apparent when perception is nonphonemic. They may also play a role in perception of the difference between the well-assimilated and the poorly assimilated members of category-goodness difference contrasts.[7]

The concept of perceptual assimilation introduced in this article, like the psychoacoustic robust/fragile distinction, calls for objective defining criteria.[8] We would offer that the like-

---

[7] We thank Michael Studdert-Kennedy for suggesting the interpretation discussed in this paragraph.

[8] We thank reviewer Kim Oller for reminding us of this important need.

lihood and direction of assimilation (i.e., to which specific native phoneme[s]) of nonnative sounds should be predictable on the basis of the degree of similarity in phonetic–articulatory features between the nonnative item and the native categories. For example, although Thompson /k'/ and /q'/ are produced with non-English ejective manner, they share the feature of stop manner of articulation with English /k/, and the places of articulation for both Thompson phones occur in allophonic variants of English /k/. In contrast, Zulu /ɔ/–/ɿ/ share neither manner nor place with any English phoneme. Phonetic similarity criteria should derive from phonetic-articulatory features as established by phoneticians (e.g., Ladefoged, 1975).

Future research could assess more directly whether English-speaking adults discriminate the clicks on the basis of auditory or phonetic information. For example, Zulu but not English listeners would be expected to show such presumably speech-specific influences as trading relations between phonetic cues (e.g., Best et al., 1981) and vowel context effects (e.g., Mann & Repp, 1980). Werker and Logan's (1985) technique of determining different discrimination patterns for auditory, phonetic, and phonemic levels of speech perception could also be applied to cross-language group comparisons and to comparisons between nonassimilable versus single-category contrasts.

The current infant findings point out an important limitation of earlier findings of perceptual decline in discrimination of early discriminated nonnative contrasts at around 10–12 months of age, which Werker et al. (1981; Werker & Tees, 1984a) attributed to their phonological irrelevance in the infants' language environment. Because the clicks are phonologically irrelevant in English, and fail to occur even as allophonic variants, the maintenance of discrimination for clicks calls for a modification of their argument, as outlined in the introduction. However, on the basis of our reasoning about perceptual assimilation, our findings are viewed as compatible with Werker's more general proposal of a developmental transition from prephonemic perception of speech sounds, which may entail a perceptual focus on either their auditory or their phonetic-articulatory properties (we favor the latter possibility: Best, 1984), to phonemic perception at around 10–12 months.

Of what use would the proposed transition to phonemic perception be for the infant's acquisition of the ambient language? The perceptual reorganization at 10–12 months that Werker found closely parallels the universal milestones of beginning word comprehension and, for many infants, the first productions of words (e.g., Lenneberg, 1967; Stark, 1980; see also Ramsay, 1980, regarding language-related neuropsychological changes at this age). The prephonemic sensitivity of infants under 10–12 months of age for many nonnative contrasts is surely well suited to their ability to learn whichever language surrounds them. However, as they become attuned to the ambient language and first begin to use words, phonemic perception should presumably aid their language acquisition. If phonemic perception entails assimilation of incoming sounds to the categories employed in the native language, then it may benefit the infant by sharpening the lines of structural organization within the phonological system of their language and by helping to establish perceptual constancy among the acoustic variations of words pronounced in different contexts and by different speakers. These benefits would presumably continue to aid efficient speech perception by adults, thus accounting for their continued difficulty with discriminating nonnative sounds that are assimilated to a single native phoneme category.

References

Abramson, A. S., & Lisker, L. (1970). Discriminability along the voicing continuum: Cross-language tests. *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 569–573). Prague: Academia.

Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology. Vol. 2. Perception* (pp. 67–96). New York: Academic Press.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effect of early experience. *Child Development, 52,* 1135–1145.

Best, C. T. (1984). Discovering messages in the medium: Speech and the prelinguistic infant. In H. E. Fitzgerald, B. M. Lester, & M. W. Yogman (Eds.), *Theory and research in behavioral pediatrics* (Vol. 2, pp. 97–105). New York: Plenum.

Best, C. T., MacKain, K. S., & Strange, W. (1982, April). *Across-language study of categorical perception for semivowel and liquid contrasts.* Paper presented at 103th meeting of the Acoustical Society of America, Chicago.

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics, 29,* 191–211.

Burnham, D. K. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics, 7,* 207–240.

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

Catford, J. C., & Ladefoged, P. (1968). Practical phonetic exercises. UCLA Working Papers in Phonetics, 11, 1–22. Los Angeles: UCLA Press.

Doke, C. M. (1926). The phonetics of the Zulu language. *Bantu studies:* Vol. 2. Johannesburg: University of Witwatersrand Press.

Eilers, R. E., Gavin, W., & Oller, D. K. (1982). Cross-linguistic perception in infancy: Early effects of linguistic experience. *Journal of Child Language, 9,* 289–302.

Eilers, R. E., Gavin, W., & Wilson, W. R. (1979). Linguistic experience and phonetic perception in infancy: A cross-linguistic study. *Child Development, 50,* 14–18.

Eilers, R. E., & Minifie, F. D. (1975). Fricative discrimination in early infancy. *Journal of Speech and Hearing Research, 18,* 158–167.

Eilers, R. E., Wilson, W. R., & Moore, J. M. (1977). Development of changes in speech discrimination in infants. *Journal of Speech and Hearing Research, 20,* 766–780.

Eimas, P. D. (1978). Developmental aspects of speech perception. In R. Held, H. W. Leibowitz, & H. L. Teuber (Eds.), *Handbook of sensory physiology: Vol. 8. Perception* (pp. 357–374). Berlin: Springer-Verlag.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science, 171,* 303–306.

Goldstein, L., & Browman, K. (1986). Representation of voicing contrasts using articulatory gestures. *Journal of Phonetics, 14,* 339–342.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia, 9,* 317–323.

Haggard, M., Ambler, S., & Callow, M. (1970). Pitch of the voicing cue. *Journal of the Acoustical Society of America, 47,* 613–617.

Horowtiz, F. D. (1975). Visual attention, auditory stimulation, and language discrimination in infants. *Monographs of the Society for Research in Child Development, 39,* 5–6, Serial No. 158.

Jusczyk, P. W. (1982). Auditory versus phonetic coding of speech signals during infancy. In J. Mehler, E. C. Walker, & M. Garrett (Eds.), *Perspectives on mental representation* (pp. 361–387). Hillsdale, NJ: Erlbaum.

Jusczyk, P. W. (1984). On characterizing the development of speech perception. In J. Mehler & R. Fox (Eds.), *Neonate cognition: Beyond the blooming, buzzing confusion* (pp. 199–231). Hillsdale, NJ: Erlbaum.

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

Ladefoged, P. (1971). *Preliminaries to linguistic phonetics.* Chicago: University of Chicago Press.

Ladefoged, P. (1975). *A course in phonetics.* New York: Harcourt, Brace & Jovanovich.

Ladefoged, P., & Traill, A. (1984). Linguistic phonetic descriptions of clicks. *Language, 60,* 1–20.

Lasky, R. E., Syrdal-Lasky, A., & Klein, R. E. (1975). VOT discrimination by four- to six-and-a half-month-old infants from Spanish environments. *Journal of Experimental Child Psychology, 20,* 215–225.

Lenneberg, E. H. (1967). *Biological foundations of language.* New York: Wiley.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431–461.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustic measurements. *Word, 20,* 384–422.

MacKain, K. S. (1982). On explaining the role of experience on infants' speech discrimination. *Journal of Child Language, 9,* 527–542.

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

Maddieson, I. (1984). *Patterns of sounds.* Cambridge, England: Cambridge University Press.

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on the perception of /ʃ/–/s/ distinction. *Perception & Psychophysics, 28,* 213–228.

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

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

Muller, Z. (1965). *The world's living languages.* New York: Frederick Unger.

Nyembezi, C. L. S. (1972). *Learn Zulu.* Pietermaritzburg, South Africa: Shuter & Shooter.

Oller, D. K., & Eilers, R. E. (1983). Speech identification in Spanish- and English-learning 2-year-olds. *Journal of Speech and Hearing Research, 26,* 50–53.

Pisoni, D. B. (1973). Auditory and phonetic codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13,* 253–260.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 297–314.

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America, 55,* 328–333.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics, 15,* 285–290.

Polka, K. (1987, May). *Perception of Persian uvular and velar stops by speakers of American English.* Paper presented at 113th meeting of the Acoustical Society of America, Indianapolis.

Ramsay, D. S. (1980). Beginnings of bimanual handedness and speech in infants. *Infant Behavior and Development, 3,* 67–77.

Signal Technology, Inc. (1986). Interactive Laboratory Systems: ILS programmer's guide. Goleta, CA: Signal Technology Inc.

Singh, S., & Black, J. W. (1966). Study of twenty-six intervocalic consonants as spoken and recognized by four language groups. *Journal of the Acoustical Society of America, 65 (Suppl. 1),* SS11.

Stark, R. E. (1980). Stages of speech development in the first year of life. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology: Vol. 1.* Production (pp. 73–92). New York: Academic Press.

Streeter, L. A. (1976). Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature, 259,* 39–41.

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

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

Weinreich, U. (1953). *Languages in contact.* New York: Linguistic Circle of New York.

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

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

Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology, 37,* 278–286.

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

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

Whalen, D. H., & Liberman, A. M. (1987). Speech perception takes precedence over nonspeech perception. *Science, 237,* 169–171.

Ziervogel, D., Louw, J. A., & Taljaard, P. C. (1976). *A handbook of the Zulu language.* Pretoria, South Africa: J. L. van Schaik.

Received April 28, 1987
Revision received August 24, 1987
Accepted August 24, 1987 ■