Divergent Developmental Patterns for Infants' Perception of Two Non-Native Consonant Contrasts*

Catherine T. Best,† Gerald W. McRoberts,‡ Rosemarie LaFleur,††† and Jean Silver-Isenstadt†††

Young infants discriminate non-native and native consonant contrasts, yet 10-12 month olds discriminate most non-native contrasts poorly, like adults. However, English-speaking adults and 6-14 month infants discriminate Zulu clicks, consistent with a model predicting that listeners who have a native phonology assimilate non-native consonants to native categories when possible, but hear Non-Assimilable (NA) consonants as nonspeech sounds (Best, McRoberts & Sithole, 1988). NA contrasts thus avoid language-specific effects and are discriminated, whereas consonants assimilated equally into a Single Category (SC) are discriminated poorly by listeners showing language-specific influences; other possible assimilation patterns show poor to excellent discrimination. This study directly compared discrimination of NA clicks and SC ejectives by 6-8 and 10-12 month olds with a conditioned fixation habituation procedure. Consistent with predictions, the younger group discriminated both non-native contrasts and a control English contrast, whereas the older group discriminated only the NA and English contrasts.

INTRODUCTION

The influence of language experience on speech perception is evident in the limitations that have been observed in adults' categorization and discrimination performance with phonetic distinctions that do not contrast phonologically in their own language(s) (e.g., Best & Strange, 1992; Flege, 1989; Flege & Eefting, 1987; Lisker & Abramson, 1970; MacKain, Best, & Strange, 1981; Miyawaki et al., 1975; Polka, 1991, 1992; Tees & Werker, 1984; Trehub, 1976; Werker & Logan, 1985; Werker & Tees, 1984a). Yet given that infants learn whichever language is used in their homes within their first few years, they obviously must be able, from fairly early on, to perceive virtually the full range of phonetic contrasts used in any of the world's languages. Research with infants under about 6 months has borne out this near-universal phonetic sensitivity for consonant and vowel contrasts. Such young infants discriminate segmental contrasts regardless of whether they occur in the native language or only in unfamiliar languages (e.g., Eilers & Minifie, 1975; Eilers, Wilson, & Moore, 1977; Eimas, 1975; Eimas & Miller, 1980; Eimas, Siqueland, Jusczyk & Vigorito, 1971; Jusczyk & Thompson, 1978; Lasky, Syrdal-Lasky, & Klein, 1975; Streeter, 1976; Swoboda, Morse & Leavitt, 1976; Swoboda, Kaas, Morse & Leavitt, 1978; Trehub, 1973, 1976). This striking developmental difference in perception of non-native phonetic contrasts indicates that sometime between early infancy and adulthood the listener's experience with a particular language comes to exert a powerful influence on speech perception.
An important theoretical issue for developmental cross-language research to resolve is the nature of this language-specific effect: When and why does the ambient language begin to leave its mark on speech perception, particularly on the perception of non-native sound patterns? Regarding the first part of the question, a number of studies indicate that language-specific perceptual effects appear before the end of the infant's first year. A possible clue to the second part is that the timing of these early perceptual changes varies for different aspects of sound patterning in speech (for in-depth discussions of possible developmental accounts, see Best, in press; Jusczyk, 1993, in press; Werker & Pegg, 1992). To summarize, Werker and her colleagues have provided strong evidence that a native-language effect on perception of consonant contrasts becomes established between 8 and 10 months of age. After 10 months, English-learning infants no longer discriminate several non-native consonant contrasts from the Hindi and Nthlakampx (a Salish language [Thompson] of the Canadian Pacific region) languages which they can clearly discriminate prior to 8 months (Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 1984b; Werker & Lalonde, 1988). Certain language-specific effects appear even earlier for vowels. English and Swedish 6 month olds each show internally-organized perceptual categories only for the vowel categories of their own language, i.e., poor discrimination of “good” tokens in the neighborhood of the category prototype but relatively better discrimination among “poor” tokens in the category periphery (Kuhl et al., 1992). And although English-learning 4-1/2 month olds can discriminate two non-English vowel contrasts from German, 6-8 month olds show a “magnet effect” asymmetry in discriminating the less vs. more English-like vowels in these German contrasts, and 10-12 month olds fail to discriminate them altogether, in either direction (Polka & Werker, 1994). Infants' attunement to some more global prosodic properties of speech may be evident even earlier than that for vowels (Mehler et al., 1988); nonetheless their attunement to certain other prosodic properties may not appear until their second half-year (Jusczyk et al., 1992). Regardless of onset age differences for these diverse aspects of speech, the findings clearly suggest that sensitivity to specific phonetic properties in speech declines if the language environment does not provide exposure to them.

This pattern of results led some to propose that exposure to specific phonetic contrasts during an early critical period is needed to maintain the neural elements that are innately tuned to the phonetic features involved, and conversely, that lack of exposure to particular contrasts results in attrition of the associated neural elements (e.g., Aslin & Pisoni, 1980; Eimas, 1975). Alternatively, it has been suggested that differential phonetic experience may sharpen attention or psychoacoustic responsiveness to phonetic properties found in the native language and/or may attenuate such responsiveness to properties that are absent from that language (e.g., Burnham, 1986; Diehl & Kluender, 1989; Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991; Pisoni, Aslin, Perey, & Hennessy, 1982; Walley, Pisoni, & Aslin, 1981). Still others have argued that, instead, differential phonetic experience shapes the higher-level processing (e.g., phonological coding, retention in memory) of auditory information from the speech signal (e.g., Tees & Werker, 1984; Werker & Tees, 1984a).

As has been pointed out elsewhere, a simple sensorineural loss explanation is untenable for several reasons (e.g., Best, 1984, 1994; MacKain, 1982; Werker & Tees, 1984a). For one, adults' perception of non-native phonetic contrasts can at least sometimes be improved by learning the other language (e.g., Flege, in press; MacKain, Best & Strange, 1981; Williams, 1979) or by laboratory training (e.g., Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991; Pisoni et al., 1982). In addition, discrimination of non-native contrasts benefits from task manipulations that reduce memory demands (e.g., Carney, Widin, & Viemeister, 1977; Werker & Logan, 1985), or that isolate the crucial acoustic cues by removing them from speech context (e.g., Miyawaki et al., 1975; Werker & Tees 1984a). Moreover, in some cases listeners have had exposure to the phonetic properties of non-native contrasts on which they have shown perceptual difficulties, because those phonetic features occur in allophonic variants of native phonological categories (e.g., MacKain, 1982). None of these observations is consistent with sensorineural attrition due to lack of exposure during an early critical period.

Of the remaining two accounts, the attentional one may be weakened, although not refuted, by reports that training or instructional manipulations which focus listeners' attention on the critical acoustic properties of non-native contrasts fail to improve perceptual performance on the associated phonetic contrasts within speech contexts (e.g., Werker & Tees, 1984a). Such findings suggest that the higher-level processing explana-
tion accounts for language-specific effects on speech perception better than does the attentional explanation. Alternatively, however, the failures may simply indicate that the attentional manipulations were inadequate for the attentional focus on the isolated cues to carry over to acoustically complex syllabic contexts.

Note that both the attentional account and the information processing account nonetheless seem to assume, explicitly or implicitly, that lack of experience with specific phonetic features or contrasts lies at the root of the ubiquitous difficulty adults have with non-native phonological contrasts. That is, they presume that phonetic experience per se is the source of the language-specific perceptual effects that emerge in infancy. This assumption was called into question by the recent finding that monolingual English-speaking adults, and infants up to the oldest age tested (14 months), showed good discrimination of click consonant contrasts from Zulu (Best, McRoberts, & Sithole, 1988). Clicks are produced by making the tongue form a complete closure around the palate (roof of the mouth), then causing a small vacuum to form by drawing the side or tip of the tongue downward. Ultimately, the tongue breaks its contact with the palate at that point and the vacuum is released, producing a suction sound or click. Click sounds fall entirely outside the range of allophonic experience with spoken English. Yet even without training and without any lowering of task demands, adults performed much better on the clicks than they had been reported to do on other non-native contrasts, whether or not the phonetic properties of those other contrasts coincide to any extent with native allophonic experience (see, e.g., Tees & Werker, 1984). Moreover, there was no developmental decline in infants' discrimination of the clicks, contrary to the marked decline at 10-12 months for discrimination of the non-native contrasts tested by Werker and colleagues.

The findings with click consonants suggest that the effect of experience on perception of non-native contrasts is not a simple effect of experience, or lack thereof, with specific phonetic features or contrasts in speech. Rather, language-specific perceptual effects must reflect listeners' knowledge of the relation between physical, phonetic properties in speech and the more abstract linguistic functions that phonological categories and contrasts serve in the native language. The phonetic properties of the other non-native contrasts tested, but not of the clicks, apparently made them susceptible to being perceived in some relation to native phonological categories. Best and colleagues (1988) posited that listeners who have become familiar with the phonological system of a specific language tend to perceptually assimilate unfamiliar non-native consonants and vowels to their own phonological categories based on phonetic similarities, if the similarities are sufficient to permit this. On the other hand, if particular non-native sound patterns deviate too greatly from the phonetic properties employed in the native phonological system—e.g., the suction-release action for the click consonants is unlike any of the phonetic features that comprise the English phonological system—listeners should fail to assimilate those sounds as potential phonological elements. In the latter event, they will instead perceive the non-assimilated speech elements as nonspeech sounds, as did the English-speaking adults who heard the Zulu clicks (Best et al., 1988).

Those suppositions form the basis of a Perceptual Assimilation Model (PAM), which is more fully described elsewhere (Best, 1993, 1994, in press a, b; Best & Strange, 1992). Its primary contribution to the empirical literature to date is that it accounts at once for both the high level of discrimination for non-native click contrasts, on the one hand, and the more commonly-reported adult perceptual difficulties and developmental decline in infant discrimination for other non-native consonant contrasts, on the other hand. The model has broader theoretical implications, however. PAM makes systematic predictions about other types of non-native contrasts, viz that discrimination levels should range from poor to excellent, depending on differences in the way the phonetic properties of non-native phonetic segments (consonants and vowels) are assimilated to native phonological categories.

To summarize, the phonetic properties of a non-native segment may bias it toward perceptual assimilation into the phonological system of the listener's native language. If assimilated into a particular native category, it may either match the ideal phonetic representation of the category, it may deviate modestly from that ideal but be heard as a good exemplar of the category, or it may fall near the category's periphery and be heard as a relatively poor exemplar of the category. Alternatively, the phonetic properties of a non-native segment may fall somewhere in between native categories, in "uncommitted phonetic space," such that it is heard as a speech sound (i.e., potential phonological element) but it is not assimilated into any specific native category. Finally, the phonetic properties of a non-native segment may be so uncharacteristic of those
employed in the native phonological system that it is not assimilated as a speech sound, but instead falls outside the phonological realm altogether and is perceived to be a nonspeech sound (i.e., environmental sound or nonlinguistic human sound such as a cough, hiss, or a disapproving "tsk-tsk"). To English speakers, the click consonants of Zulu fail to be assimilated as potential elements of a phonological system, and are heard as nonspeech (Best et al., 1988).

Predicted discrimination levels for non-native contrasts follow from the assimilation patterns of each of the contrasting segments. When two non-native segments are assimilated into a single native category, discrimination should be poor if both fall equally close to the native category ideal, a case referred to as Single Category assimilation (abbreviated SC). It has been argued that many of the non-native contrasts for which adults and older infants have been reported to show perceptual difficulties are likely to be SC contrasts (e.g., Best, 1993, 1994, in press a, b). On the other hand, when two non-native segments are assimilated into a single native category, but unequally such that one is close to the native ideal while the other is in the category periphery, listeners should perceive a Category Goodness difference (CG) between them. In CG contrasts, discrimination is relatively good, the exact level depending on the magnitude of the goodness difference between the two sounds and their proximity to the periphery of the native category. Discrimination should be even better, approaching native listener levels, when the contrasting non-native segments are assimilated to Two Categories (TC) in the native phonology. One or both of the non-native segments may instead fall in Uncommitted phonetic space (UC or UU, respectively), leading to relatively good discrimination in the first case or moderate to poor discrimination in the second. Finally, the contrasting non-native categories may be Non-Assimilable (NA) with respect to the native phonological system, as described above for the Zulu clicks, in which case they should be discriminated moderately to very well (for more detailed description, see Best, in press a, b).

A small number of studies has examined PAM's predictions for adult perception of non-native contrasts (Best et al., 1988; Best & Strange, 1992; Polka, 1991, 1992, submitted). Their findings have supported the model's predictions (all types of non-native contrasts have been tested, except those with assimilation to uncommitted phonetic space). However, extending PAM to explain language-specific developmental changes in infant speech perception is problematic at present. There has been only one published report on infants, which looked only at the NA assimilation type. Most importantly, the NA type has not been compared to an assimilation pattern for which a developmental decline in discrimination would instead have been predicted (Best et al., 1988). Moreover, comparison of the click findings to other cross-language infant studies are confounded by a difference in methodology. Whereas Werker's studies employed the conditioned head-turn response in the multi-trial go/no go procedure used at a number of infant speech perception laboratories (e.g., Eilers, Wilson, & Moore, 1977; Kuhl, 1980), Best and colleagues (1988) used a conditioned visual fixation response in an infant-controlled habituation-dishabituation procedure (Miller, 1983). The conditioned fixation procedure had not been used previously in tests of infant consonant perception. It is at least plausible that it may be cognitively less demanding and/or psychophysically more sensitive than the conditioned head-turn procedure. If so, the divergence between the 10-12 month decline in discrimination of Werker's Hindi and Nthlakampx stimuli and the lack of developmental change in discrimination of Zulu clicks might be attributable solely to the difference in methodology.

Therefore, it was important to verify the robustness of the NA pattern for click contrasts, and test the contrary prediction of developmental decline for SC contrasts, in the same infants and using a single methodology. For this purpose, we used the conditioned fixation procedure to test 6-8 month old and 10-12 month old English-learning American infants on three contrasts: native English /ba/-/da/, Nthlakampx velar vs. uvular ejectives /k'wa/-/q'wa/, and Zulu voiceless unaspirated apical vs. lateral clicks /lа/-/lа/. These ages were tested because previous reports indicated that the younger age should discriminate native and non-native consonant contrasts without difficulty, whereas the older age should show marked difficulty in discriminating non-native consonant contrasts other than the clicks. The English and Nthlakampx stimuli were those used by Janet Werker (Werker & Teens, 1984a, b). The Zulu click stimuli were those used Best and colleagues (1988). The English contrast served as a native control comparison. The clicks had met the criteria for an NA assimilation type according to the adult findings of Best et al., and had shown good discrimination across all ages, without a developmental decline in infants' discrimination. The Nthlakampx ejectives were
expected to fit the pattern for SC assimilation, whereby adult English listeners tend to assimilate both the velar and the uvular ejective as equally “odd” exemplars of the English voiceless stop /k/. The SC assimilation prediction for infants was that there should be good discrimination at 6-8 months but poor discrimination at 10-12 months. Such a pattern would replicate the Werker and Tees (1984b) findings in a different laboratory and with a different infant testing procedure.

The Nthlakampx contrast in particular was selected for several reasons. The perceptual results for both adults and older infants differ dramatically between this poorly-discriminated ejective contrast and the easily-discriminated click contrast. Nonetheless, these ejectives and the voiceless unaspirated clicks show a number of similar acoustic properties (see Best et al., 1988; Werker & Tees, 1984a). Both types of consonants produce brief noise bursts that are higher in amplitude than the following vowel, and show similar high frequency poles in the noise spectrum around 4200-4600 Hz. The noise bursts for both consonant contrasts are separated from the subsequent vowel by a brief silent interval, thus the vowel is unlikely to produce masking of the noise in either case (see Werker & Tees, 1984a). There are, of course, some acoustic differences between the ejectives and the clicks. The click noise bursts are 9 m longer ($M = 47.4$ ms) than the bursts for the ejectives ($M = 38.5$ ms). However, the total duration of click + silence ($M = 66.9$ msec) is equal to that for ejective burst + silence because the silent interval is shorter for clicks ($M = 19$ msec) than for ejective bursts ($M = 29$ ms). Thus there is slightly more likelihood of vowel masking for the clicks, which would work against good discrimination. The noise bursts differ between the two clicks much more strikingly in the frequency of the lower spectral pole (120 Hz vs. 2450 Hz) than do those of the ejectives (3100 Hz vs. 3200 Hz). The noise bursts for the two types of consonant contrasts most likely also differ strongly in their amplitude envelopes.

Method

Subjects. Twenty-four infants were included in the study, 12 at 6-8 months (9 males, 3 females; $M_{age} = 6$ mo. 18 days, range = 5 mo. 30 days to 7 mo. 26 days) and 12 at 10-12 months (5 males, 7 females; $M_{age} = 11$ mo. 7.5 days, range = 9 mo. 26 days to 12 mo. 14 days). All were normal, full-term infants without gestational or labor/delivery complications, and were free of ear infections or colds on the day of testing. An additional 39 infants were excluded from the final data set (crying = 16; equipment problems = 6; experimenter error = 3; inattentiveness = 13 [i.e., 10 or more consecutive trials without fixation responses]; Down syndrome = 1).

Stimulus materials. The three stimulus contrasts used in this study were the English /bal-/ /da/, Nthlakampx ejectives /kæ/-qær/, and Zulu clicks /la/-/ja/. All stimulus contrasts included multiple natural tokens produced by a native speaker of the language involved, selected for similarity in duration, amplitude and frequency characteristics of the tokens within the pairs of contrasting categories. The English and Nthlakampx contrasts were produced by male adult speakers; the Zulu contrast was produced by an adult female. Acoustic measurements of the non-native contrasts are reported in the original papers (Best et al., 1988; Werker & Tees, 1984a, b).

Procedure. We employed the same conditioned visual fixation habituation procedure used in our previous study (Best et al., 1988; see also Miller, 1983). In this procedure, tokens of one speech category were played to the infant over a hidden loudspeaker at a conversational listening level whenever the infant fixated on a rear-projected picture (colored checkerboard) presented on screening material affixed to a window in the wall they face during testing. A video monitor connected to a hidden camera at the side of the projection window displayed the infant’s head and shoulders to an observer in the adjacent room, who registered the infant’s fixations (as well as bouts of crying and sleeping) via key press input to a computer (Atari 800). The computer registered the fixation duration, and controlled the presentation of audio stimuli from a reel-to-reel tape deck (Otari 5050 MXB). When the infant looked away from the picture, the observer released the “looking” key and the computer stopped the presentation of the speech sounds to the infant.

Trial duration was under infant control: if the infant looked away from the slide for two consecutive seconds the trial ended and the slide blankened during the intertrial interval (ITI). After one second the slide automatically reappeared, beginning the next trial. Habituation was defined as two consecutive trials with fixation durations below 50% of the mean of the two highest preceding trials (Miller, 1983). The criterion was calculated and updated on a trial-to-trial basis by the computer program. Once habituation was met during the first phase,
referred to as the familiarization phase, audio presentations shifted to the contrasting speech category for the test phase, which continued until the infant again habituated. The index of discrimination is any change in fixation during the first two test trials relative to the last two familiarization trials. Full technical details for the procedure are available in the original report (Best et al., 1988).

During testing the infant sat in an infant seat or on the parent’s lap, about 3 feet from the rear-projection window, in the dimly lit testing room. Both were seated in a small booth constructed by attaching two partitions to the wall on either side of the projection window, about 3.5 ft. apart. Each partition was approximately 6 ft. high, and extended 4 ft out from the wall. The booth was open at the back, and its sides were covered with black fabric. The wall at the front of the booth was also covered with black fabric, except for the 2 ft. × 2 ft. area directly in front of the infant’s head where the picture was projected. A Jamo loudspeaker was used for stimulus presentations; it was attached to the wall 3 ft. above the projection window, and was hidden behind the black cloth covering the wall. Speech was presented to the infant at a 65-70 dB sound pressure level. Both the parent and the infant observer listened to music over closed-design headphones (Sennheiser HD440) during testing to prevent them from inadvertently biasing the infant’s behavior or the fixation observations.

Each infant completed all three speech discrimination tests within a single session. Test order was randomized across infants within each age group. Short breaks of 5-10 minutes were taken between tests if necessary to maintain infants’ attention and/or to soothe them if they had become irritable. Otherwise, the session proceeded from one test to the next with only the 1-2 minute break needed for re-positioning the audio tape and restarting the computer program. Infants were eliminated from the final data set if they cried for more than a cumulative 30 seconds during any test, or if they cried during any of the trials just before or after the test shift.

Results

Inter-observer reliability. The data for a random selection of seven of the infant subjects (i.e., 21 individual tests) were rescored by second observers re-running the testing program while viewing the videotapes. Thus, interobserver reliability was assessed off-line. Reliability was quite good (r = .91 to .985).

Habituation during Familiarization. Habituation during the familiarization phase of the tests was verified by analyses of variance (ANOVA) on both forward and backward habituation curves. Forward habituation analyses compared the mean fixation in the first two trials against the mean in the final two trials prior to the stimulus shift in all tests, in an Age (2) × Language (3) × Trial Block (2) ANOVA. The Trial Block effect revealed a significant decline in fixation from the first familiarization trials (M = 12.36 s) to the final trials before the test shift (M = 2.46 s), F(1, 22) = 87.476, p < .0001. The Age effect was also significant, F(1, 22) = 6.081, p < .025, indicating that the younger group looked significantly longer during familiarization (M = 9.02 s) than did the older group (M = 5.8 s). However, an Age × Trial Block interaction showed this age difference to be restricted to the beginning trials of the familiarization phase (M = 15.48 and 9.25 s, respectively); both ages habituated to the same low fixation level by the final preshift trials (M = 2.56 and 2.36 s, respectively), F(1, 22) = 8.104, p < .01.

A separate Age (2) × Language (3) × Trials (4) ANOVA was conducted on the backward habituation data, for the last four trials of familiarization. The Trials effect showed a dramatic and significant decline in fixations during the last two preshift trials (trials -2 and -1: M = 2.67 and 2.25 s) relative to the two trials just preceding those (trials -4 and -3: M = 11.27 and 13.45 s), F(3, 66) = 18.313, p < .0001 (see Figure 1). A significant Age effect, F(1, 22) = 5.68, p < .03, indicated that the younger infants fixated longer during these familiarization trials (M = 9.06 s) than did the older infants (M = 5.76 s). This age difference was evident only during the -4 and -3 trials before the shift to the test phase (6-8 month M = 13.14 and 18.0 s; 10-12 month M = 9.40 and 8.91 s). As the forward habituation analysis had shown, fixation had dissipated to the same low fixation levels for both ages by the two trials just preceding the shift (6-8 month M = 2.97 and 2.20 s; 10-12 month M = 2.42 and 2.30 s).

Given that both forward and backward habituation values are constrained by the habituation criteria we used, we also examined possible language and age differences in the number of trials to habituation, and in mean looking time per habituation trial in separate Age (2) × Language (3) ANOVAs. These two indices are not constrained by the method we used for determining habituation. No main effects or interactions approached significance in either of the latter ANOVAs.
Infant Perception of Non-Native Contrasts

Figure 1. Backward habituation curves for the last 4 trials of the familiarization phase, with standard error bars, shown separately for 6-8 and 10-12 month olds. For comparison, the mean fixation time for the first two trials and the mean for the two highest trials are also displayed.

**Discrimination results.** Discrimination was assessed by comparing the mean fixation duration during the last two trials of the familiarization phase (Preshift Trial Block) against the mean fixation during the first two trials of the test phase (Postshift Trial Block). The postshift block was defined as beginning with the first trial after the stimulus shift in which the infant fixated on the slide and thus had an opportunity to begin hearing the test stimuli (see Best et al., 1988). A significant increase in fixation during the postshift block relative to the preshift block is taken to indicate that infants detected the stimulus change. These data were entered into Language (3) x Trial Block (Preshift vs. Postshift) ANOVAs; test order was left out as a factor because preliminary analyses showed that it did not have any systematic effect on discrimination. Separate ANOVAs were conducted for each age to test the a priori predictions that 6-8 month olds would discriminate all three contrasts whereas 10-12 month olds would discriminate only the English and Zulu contrasts, and would fail on the Nthlakampx contrast. Given these predictions, simple effects tests of the Language x Trial Block interaction for each age were also carried out as planned comparisons.

**6-8 month olds.** This group's main effect for Trial Block, $F(1, 22) = 17.02, p < .002$, revealed significant recovery of fixation overall during the postshift trials ($M = 6.24$ s) relative to the preshift trials ($M = 2.56$ s), as expected. In addition, there was a significant Language effect, $F(2, 22) = 4.16, p < .03$. According to Tukey tests, this effect is attributable to significantly greater fixation for the English test ($M = 6.33$ s) than for the
Nthlakampx test ($M = 3.27$ s), $p < .05$, during the trials surrounding the shift. Although the Language $\times$ Trial Block interaction was not itself significant, a simple effects test on the interaction was conducted, as planned, to determine whether recovery of fixation during the initial test trial block was significant for each of the three contrasts. The results supported predictions. This age group showed significant recovery of fixation on the initial test trials for all three languages: English, $F(1, 11) = 7.09, p < .025$; Zulu, $F(1, 11) = 4.87, p < .05$; and Nthlakampx, $F(1, 11) = 4.67, p = .05$. The simple effects tests also suggested that the main effect for Language could be traced primarily to fixation differences during the test trials, $F(2, 11) = 2.97, p = .07$, rather than the preshift trials, $ns$. Postshift fixation was much higher for English ($M = 9.19$ s) than for Nthlakampx ($M = 4.52$ s), whereas preshift fixation was more nearly equivalent ($M s = 3.5$ vs. 2.03 s, respectively). Thus, while these infants showed significant discrimination on both tests, there is the suggestion of a mild language-specific effect in degree of discrimination for these two languages.

10-12 month olds. This age group also showed a significant Trial Block effect, $F(1, 11) = 11.86, p < .004$, indicating overall discrimination (preshift $M = 2.36$ s; postshift $M = 5.09$ s). In contrast to the younger group, the Language effect was nonsignificant, while the Language $\times$ Trial Block interaction was marginally significant, $F(2, 22) = 2.67, p = .09$. The planned simple effects test on this interaction revealed, as expected, that the older infants showed significant recovery of fixation only for English, $F(1, 11) = 10.53, p < .008$, and for Zulu, $F(1, 11) = 5.69, p < .04$, but not for Nthlakampx, $ns$. A comparison of the discrimination results for the two ages is shown in Figure 2.

![Figure 2](image-url)
DISCUSSION

The findings strongly support the prediction of diverging developmental patterns for infants' perception of the two non-native consonant contrasts tested. The younger infants discriminated both non-native contrasts and, of course, the native English contrast. The older infants discriminated not only the English contrast but also the Zulu clicks, consistent with predictions for a NA contrast (Non-Assimilable) according to the Perceptual Assimilation Model (PAM), yet they failed to discriminate the Nthlakampx ejectives, consistent with predictions for SC assimilation (Single Category). The present study directly compared, in a single within-subjects investigation, two important but disparate cross-language speech perception findings with infants. Werker and Tries' (1984b) finding of developmental decline between 6-8 months and 10-12 months for English-learning infants' discrimination of the /k/-q/ ejective distinction has now been replicated in an independent laboratory, and has been extended to a different methodological technique. We also replicated, in the same sample of infants, the earlier report that English-learning infants nonetheless continue to discriminate the /a/-/ja/ click contrast even at 10-12 months, a finding which stands at odds with reports on perception of other non-native consonant contrasts by that age group.

The divergent developmental pattern for these two contrasts cannot be explained by any difference in phonetic exposure, since neither ejectives nor clicks occur as allophones of any English consonants. For this reason (as well as those provided in the Introduction), neither can the discrepant developmental trends for these two non-native contrasts be explained by differences in neural attrition due to differential phonetic exposure. Nor could the click vs. ejective discrimination difference at 10-12 months be caused by different degrees of auditory masking. The bursts of both the clicks and the ejectives, which appear to carry the primary information about both of these place of articulation contrasts, are separated from the following vowel by a silent gap that should sufficiently attenuate potential masking of the burst by the vowel.

Some might, alternatively, posit a difference in acoustic salience as an explanation of the difference in older infants' discrimination of the clicks versus the ejectives (see Burnham, 1986). However, as noted in Best et al. (1988), no objective criteria for salience have been proposed that are independent of the discrimination levels that salience is supposed to explain, thus the concept is tautological. In the present study, differential acoustic salience would be difficult to argue in any event, because these two particular non-native contrasts are quite similar in their acoustic properties, as described in the Introduction.

We suggest, instead, that the answer can be found in the development of perceptual ability to relate the physical, phonetic properties of speech to the more abstract phonological categories of the native language, as posited by PAM (Best, 1993, 1994, in press a, b; Best et al., 1988). Very young infants would not yet be expected to have determined the patterning of the native phonological system, and so it should not be surprising that they perceive the phonetic properties of both native and non-native segmental contrasts. However, at least by sometime in the second half of their first year, infants begin to recognize certain basic characteristics of the native phonological system, which in turn begins to influence their perception of non-native segmental contrasts. The present findings are consistent with the notion that discrimination of a non-native consonant contrast will be retained even at 10-12 months if the phonetic properties of that contrast place it outside the general patterning of phonetic properties within the native phonological system, that is, if the non-native consonants fit the definition of a NA contrast. Such was the case here for the Zulu click consonants. If on the other hand the non-native consonants fit the definition of a Single Category assimilation type, older infants would be expected to perceive them as phonetically equivalent and essentially indistinguishable, once they have begun to recognize certain basic properties of the native phonological system. The discrimination of the Nthlakampx ejectives by 6-8 month olds, but not by 10-12 month olds, fits this prediction as well. At the present time it is unclear, however, whether infants actually assimilate non-native consonants into native phonological categories in the same way that adults do, or even whether they have yet fully-specified phonological categories (see also Werker & Pegg, 1992). It would be reasonable to suppose that the answer to those two questions is "no," given that several years of further phonological and linguistic development must still take place after the first birthday before children have achieved adult-like levels of language competence. Additional research will be needed to address the issues of non-native phonetic assimilation and development of phonological categories in infants.
Interestingly, the data gathered in the present study with the conditioned fixation procedure revealed a hint of a language-specific effect in perception of a non-native contrast even at 6-8 months. That age showed more attention to English around the shift (primarily a difference in recovery of fixation at the stimulus shift) than to Nthlakampx velar-uvular ejectives, the latter being the same contrast that shows a significant decline in discrimination just a few months later. This language-specific bias was evident in the 6-8 month olds even though they nonetheless showed significant discrimination of the Nthlakampx ejectives. The pattern of language-specific change in perception of these ejectives differs strongly from the lack of developmental change in discrimination of the Zulu clicks. We suggest that this difference is due to perceived similarities between the ejectives and native voiceless stop consonants, but a lack of perceived similarities between the clicks and any English consonants. Again, however, further work will be needed to verify whether infants do indeed perceive phonetic similarity between ejectives and voiceless stops at the same place of articulation.

Given the suggestion by some (e.g., Burnham, 1968) that contrasts which remain easily discriminated, even without phonetic exposure may be acoustically salient, several characteristics of click consonants are of special note. If clicks are acoustically salient, they should presumably be easy to perceive and/or produce as phonological elements in the languages that employ them. In addition, they should be widespread across languages. But in fact, clicks are quite rare, being found only in the Khoisan languages of Africa, the language family that is the origin of the click consonants, and in those Bantu languages which borrowed the clicks from Khoisan-speaking groups over centuries of frequent interaction (Herbert, 1990). Linguists have posited a correlation between the ease of perceiving and/or producing a phonetic contrast and the commonness of that contrast across languages (e.g., see Lindblom, Krull, & Stark, 1993; Lindblom & Maddieson, 1988). Given the rareness of clicks across languages, they should be relatively difficult to discriminate when perceived as phonological elements. In keeping with this prediction, it is claimed that "to the untrained ear there is much confusion within the class" of click consonants (Herbert, 1990, p. 123) when non-click languages borrow words from a click language for example, the "borrowing" Bantu languages typically conflated a number of the original click distinctions found in the originating Khoisan languages, thus ending up with many fewer click distinctions than existed in the target language (Herbert, 1990). Additional evidence suggests articulatory difficulties with the production of clicks as phonological elements in languages. After the particular apical versus lateral click contrast we examined here, historical evidence indicates a strong tendency for those clicks to be conflated with others from the Khoisan languages. Specifically, the lateral click is currently disappearing in Zulu. The apical is next most likely to disappear in the adopting Bantu languages, such that the palatal has become the sole click in languages such as Sesotho (Herbert, 1990). In addition, anecdotal evidence (and a small amount of systematic observational evidence) on development indicates that the clicks develop relatively late in native-speaking children's productions (Jakobson, 1958; Louw, 1964). As was the case with click-borrowing languages, young children learning click languages show a strong preference to substitute palatal clicks in place of the lateral and apical clicks (Connelly, 1984; Herbert, 1983).

The perceptual findings from the present study are consistent with predictions based on PAM that there should be divergent developmental paths for perception of different types of non-native consonant contrasts. The present study supported the hypothesis that discrimination would show a developmental decline for a non-native contrast that adults are likely to assimilate into a single category in their native phonology. Complementary to that developmental pattern, support was also found for the prediction that discrimination would remain high across development for a contrast that adults fail to assimilate to any native categories, and therefore hear as nonspeech sounds. Further research will be needed, however, to corroborate PAM's predictions about other types of assimilation patterns, including TC (Two Category) and CG (Category Goodness difference) assimilation types.

REFERENCES


of Experimental Psychology: Human Perception and Performance, 8, 297-314.

**FOOTNOTES**

* *Infant Behavior & Development*, in press.
1 Also Wesleyan University, Department of Psychology, Middletown.
2 Stanford University, Department of Psychology, Palo Alto.
3 Wesleyan University, Department of Psychology, Middletown.