INFANT BEHAVIOR AND DEVELOPMENT 18, 339–350 (1995)

981

# Divergent Developmental Patterns for Infants' Perception of Two Nonnative Consonant Contrasts

CATHERINE T. BEST
*Wesleyan University* and *Haskins Laboratories*
GERALD W. MCROBERTS
*Stanford University*
ROSEMARIE LAFLEUR AND JEAN SILVER-ISENSTADT
*Wesleyan University*

Young infants discriminate nonnative and native consonant contrasts, yet 10–12-month-olds discriminate most nonnative contrasts poorly, like adults. However, English-speaking adults and 6–14 month-old infants discriminate Zulu clicks, consistent with a model predicting that listeners who have a native phonology assimilate nonnative consonants to native categories when possible but hear non-assimilable (NA) consonants as nonspeech sounds (Best, McRoberts, & Sithole, 1988). Non-assimilable 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 nonnative contrasts and a control English contrast, whereas the older group discriminated only the NA and English contrasts.

speech perception native contrasts nonnative contrasts native categories discrimination developmental patterns

The influence of language experience on speech perception is evident in limitations in adults' categorization and discrimination of 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, 1984b). Yet given that infants learn whichever language is used in their homes within their first few years, they obviously must be able to perceive, from fairly early on, 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, Kaas, Morse, & Leavitt, 1978; Swoboda, Morse, & Leavitt, 1976; Trehub, 1973, 1976). This striking developmental difference 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 is the nature of this language-specific effect: When and how

---

We thank the Wesleyan students who helped collect data for this study: Shama Chaiken, Laura Klatt, Pam Spiegel, Amy Wolf, David Fleishman, Leslie Turner, Ashley Prince, Ritaelena Mangano, Pia Marinangelli, Heidi Queen, Peter Kim, and Jane Womer. We also thank Janet Werker for the loan of her Nthlakampx (Thompson Salish) and English stimulus materials and for numerous helpful discussions about the findings. We are especially grateful to the parents of our infant subjects for their participation. This work was supported by NIH Grants HD01994 to Haskins Laboratories and DC00403 to C. T. B.; completion of the final draft was partially supported by NATO Grant CRG 931533 to C.T.B. The research was begun when G.W.M. was a doctoral student at the University of Connecticut and a research assistant at Haskins Laboratories.

Correspondence and requests for reprints should be sent to Catherine T. Best, Haskins Laboratories, 270 Crown Street, New Haven, CT 06511.

does the ambient language begin to leave its mark on speech perception, particularly the perception of nonnative 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 1st 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, see Best, 1994a, 1994b; 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 nonnative consonant contrasts from Hindi and Nthlakampx (a Salish language [Thompson] of the Canadian Pacific region) which they can clearly discriminate prior to 8 months (Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Lalonde, 1988; Werker & Tees, 1984a). Certain language-specific effects may appear even earlier for vowels. English and Swedish 6-month-olds each show internally organized perceptual categories only for the vowels of their own language, that is, poor discrimination of "good" tokens in the neighborhood of the category prototype but relatively better discrimination among "poor" tokens in the category periphery (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). And whereas English-learning 4½-month-olds can discriminate two non-English vowel contrasts from German, 6–8-month-olds show an asymmetry in discriminating the less versus more English-like vowels in these German contrasts, and 10–12-month-olds fail to discriminate them altogether (Polka & Werker, 1994). Infants' attunement to some more global prosodic properties of native speech may be evident even earlier than that for vowels (e.g., Mehler et al., 1988); nonetheless, their attunement to certain other prosodic properties of the ambient language may not appear until the 2nd half-year (e.g., Jusczyk et al., 1992). Regardless of onset-age differences for experience-related developmental change in perception of 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, & Hennessey, 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, 1984b). As another alternative, the developmental change in perception of nonnative speech distinctions may reflect the infant's transition from a prelinguistic to a truly linguistic stage of processing speech (e.g., Jusczyk, 1992, in press; Werker & Pegg, 1992).

As has been pointed out elsewhere, a simple sensorineural loss explanation is untenable (e.g., Best, 1984, 1994a; MacKain, 1982; Werker & Tees, 1984b). Counterevidence includes findings that adults' perception of nonnative phonetic contrasts can improve with learning the other language (e.g., Flege, in press; MacKain et al., 1981; Williams, 1979) or with laboratory training (e.g., Lively et al., 1993; Logan et al., 1991; Pisoni et al., 1982). In addition, discrimination of nonnative 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 1984b). Moreover, in some cases, listeners have had exposure to the phonetic properties of nonnative 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 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 nonnative contrasts fail to improve perceptual performance on the associated phonetic contrasts within speech contexts (e.g., Werker & Tees, 1984b). Such findings imply that the higher level processing explanation, or the linguistic-processing explanation, may account 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.

Both the attentional and the information-processing accounts, like the sensorineural account, assume that lack of experience with specific phonetic features or contrasts lies at the root of adults' difficulty with nonnative phonological contrasts. 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 et al., 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 nonnative contrasts, whether or not the phonetic properties of those other contrasts coincide to any extent with native allophonic experience (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 nonnative contrasts tested by Werker and colleagues.

The findings with click consonants suggest that language-specific influences on perception of nonnative contrasts are not a simple effect of exposure, or lack thereof, to 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. That is, the effects appear to reflect linguistic processes. The phonetic properties of the other nonnative 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 nonnative 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 nonnative 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) those authors proposed that listeners should fail to assimilate those sounds as potential phonological elements. In the latter event, the nonassimilated speech elements would be perceived as nonspeech sounds, as was indeed reported by 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, 1994a, 1994b, in press; Best & Strange, 1992). Its primary contribution is that it accounts at once for both the high level of discrimination for nonnative click contrasts, on the one hand, and the more commonly reported adult perceptual difficulties and developmental decline in infant discrimination for other nonnative consonant contrasts, on the other hand. The model has broader theoretical implications, however. PAM makes systematic predictions about other types of nonnative contrasts, namely, that discrimination levels should range from poor to excellent, depending on differences in the way the phonetic properties of nonnative phonetic segments (consonants and vowels) are assimilated to native phonological categories.

To summarize, the phonetic properties of a nonnative 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 nonnative 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 nonnative 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 reported to sound like nonspeech events (Best et al., 1988).

Predicted discrimination levels for nonnative contrasts follow from the assimilation patterns of each of the contrasting segments. In adults, assimilation patterns can be assessed from their phonetic categorizations or descriptions of the nonnative elements, and their goodness-of-fit ratings for the categories they have indicated. Actual assimilation patterns are more difficult to establish in infants, but it may be possible to tap them indirectly by testing for perceptual prototype effects, within-category perceptual equivalencies, and/or ease of category formation. At least for adults, when two nonnative segments are assimilated into a single native category, PAM predicts that discrimination should be poor if both segments fall equally close to the native category ideal, a case referred to as single category (SC) assimilation. It has been argued that many of the nonnative contrasts for which adults and older infants have been reported to show perceptual difficulties are likely to be SC contrasts (e.g., Best, 1993, 1994a, 1994b, in press). On the other hand, when two nonnative segments are assimilated into a single native category, but unequally such that one is close to the native ideal whereas the other is in the category periphery, listeners should perceive a category goodness (CG) difference between them. In CG contrasts, discrimination is relatively good; the exact level depends 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 nonnative segments are assimilated to two categories (TC) in the native phonology. One or both of the nonnative 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 nonnative categories may be non-assimilable (NA) with respect to the native phonological system, as described previously for the Zulu clicks, in which case they should be discriminated moderately to very well (for more detailed description, see Best, 1994b, in press).

A small number of studies has examined PAM's predictions for adult perception of nonnative contrasts (Best et al., 1988; Best & Strange, 1992; Polka, 1991, 1992, 1995). Their findings have supported the model's predictions (all types of nonnative 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 a contrast that fits the NA assimilation type for adults. Most importantly, this contrast has not been compared to a different type for which a developmental decline in discrimination would instead be 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 multitrial go/no go procedure used at a number of infant speech perception laboratories (e.g., Eilers et al., 1977; Kuhl, 1980), Best and colleagues (1988) used a conditioned visual fixation response in an infant-controlled habituation-dishabituation procedure (first used for speech perception by 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 developmental pattern for click contrasts and to test the same infants, using the same methodology, on the contrary prediction of developmental decline for contrasts such as those used by Werker and colleagues, which are expected to be assimilated by English-speaking adults according to the SC pattern. 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 versus uvular ejectives */k'æ/–/q'æ/*, and Zulu voiceless unaspirated apical versus lateral clicks */ǀa/–/ǁa/*. These ages were tested because previous reports indicated that the younger age should discriminate native and nonnative consonant contrasts without difficulty, whereas the older age should show marked difficulty in discriminating nonnative consonant contrasts other than the clicks. The English and Nthlakampx stimuli were those used by Janet Werker (Werker & Tees, 1984a, 1984b). The Zulu click stimuli were those used by Best et al. (1988). The English contrast served as a native control comparison. The clicks had met the criteria for a 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 corresponding 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 (1984a) 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 (Best et al., 1988; Werker & Tees, 1984b). Both types of consonants produce brief noise bursts which 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 (Werker & Tees, 1984b). There are, of course, some acoustic differences between the ejectives and the clicks. The click noise bursts are 9 ms longer ($M$ = 47.4 ms, range = 36–54ms) than the bursts ($M$ = 38.5 ms, range = 35–41ms) for the ejectives. However, the total duration of click + silence ($M$ = 66.9 ms, range = 36–91 ms) is equal to that for ejective burst + silence, because the silent interval is shorter for clicks ($M$ = 19 ms, range = 0–37 ms) than for ejective bursts ($M$ = 29 ms, range = 25–35 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; $M$ = 6 months, 18 days, range = 5 months, 30 days to 7 months, 26 days) and 12 at 10–12 months (5 males, 7 females; $M$ = 11 months, 7.5 days, range = 9 months, 26 days to 12 months, 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 /ba/–/da/, Nthlakampx ejectives /k'æ/–q'æ/, and Zulu clicks /la/–/lla/. 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 nonnative contrasts are reported in the original papers (Best et al., 1988; Werker & Tees, 1984a, 1984b).

## 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 faced 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 according to corneal reflection of the visual target on the infant's pupil (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 2 consecutive s, the trial ended and the slide went blank during the intertrial interval (ITI). After 1 s, 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 ft (0.91 m) 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 (1.07 m) apart. Each partition was approximately 6 ft (1.83 m) high and extended 4 ft (1.22 m) 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 (0.61 m × 0.61 m) 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 (0.91 m) above the projection window and was hidden behind the black cloth covering the wall. Speech was presented to the infant at a 65- to 70-dB sound pressure level. Both the parent and the 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 to 10 min 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- to 2-min break needed for repositioning 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 s during any test, or if they cried during any of the trials just before or after the test shift.

# RESULTS

## Interobserver Reliability

The data for a random selection of 7 of the infant subjects (i.e., 21 individual tests) were rescored by second observers rerunning the testing program while viewing the videotapes. Thus, interobserver reliability was assessed offline; it was evaluated statistically via rank-order correlations of the per-trial looking times registered by the first and second observers. Reliability was quite good ($r = .910–.985$).

## Habituation During Familiarization

Habituation during the familiarization phase of the tests was verified by analyses of variance (ANOVAs) on both forward and backward habituation curves, as well as by ANOVAs on number of trials required to reach the habituation criterion and the mean look duration per trial during the habituation phase of each test. 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), $F(1, 22) = 87.48$, $p < .0001$. The age effect was also significant, $F(1, 22) = 6.08$, $p < .025$, indicating that the younger group looked significantly longer during familiarization ($M$ = 9.02 s) than did the older group ($M$ = 5.80 s). However, an age × trial block interaction showed this age difference to be restricted to the beginning trials of the familiarization phase ($Ms$ = 15.48 and 9.25 s, respectively); both ages habituated to the same low fixation level by the final preshift trials ($Ms$ = 2.56 and 2.36 s, respectively), $F(1, 22) = 8.10$, $p < .01$.

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: $Ms$ = 2.67 and 2.25 s, respectively) relative to the two trials just preceding those (Trials −4 and −3: $Ms$ = 11.27 and 13.45 s, respectively), $F(3, 66) = 18.31$, $p < .0001$ (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 $Ms$ = 13.14 and 18.00 s, respectively; 10–12-month $Ms$ = 9.40 and 8.91 s, respectively). 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 $Ms$ = 2.97 and 2.20 s, respectively; 10–12-month $Ms$ = 2.42 and 2.30 s, respectively).

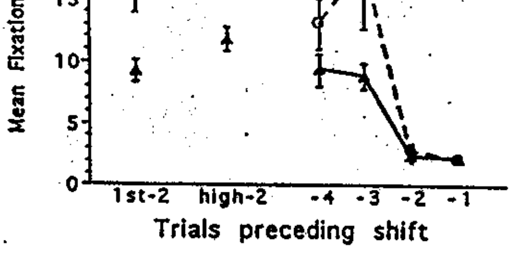

**Figure 1.** Backward habituation curves for the last four trials of the familiarization phase, with standard error bars, shown separately for 6–8- and 10–12-month-olds. For comparison, the mean fixation time for the first two trials and the mean for the two highest trials are also displayed.

Given that both forward and backward habituation values are constrained by the habituation criterion 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.

## 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 (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) × 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 à 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 × 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 × 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 = 3.50 vs. 2.03 s, respectively). Thus, although 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, whereas the language × 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.

## DISCUSSION

The findings strongly support the prediction of diverging developmental patterns for infants' perception of the two nonnative consonant contrasts tested. The younger infants discriminated both nonnative 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 according to the Perceptual Assimilation Model (PAM), yet they failed to discriminate the Nthlakampx ejectives, consistent with predictions for SC assimilation. This study directly compared, in a single within-subjects investigation, two important but disparate cross-language speech perception findings with infants. Werker and Tees' (1984a) 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 /ǀa/–/ǁa/ click contrast even at 10–12 months, a finding which stands at odds with reports on perception of other nonnative consonant contrasts by that age group.

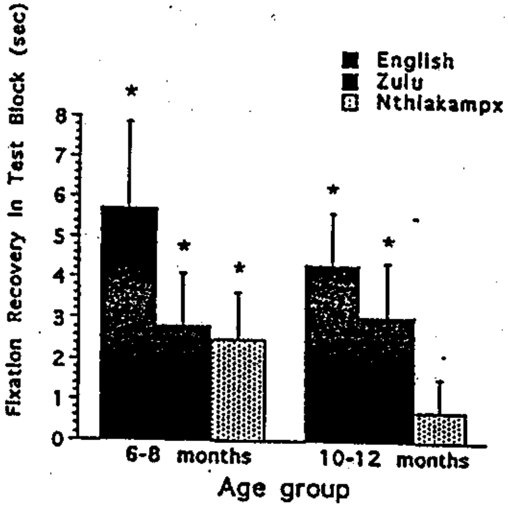

**Figure 2. Recovery of fixation responses displayed as difference scores (preshift trial block – postshift trial block), with standard error bars, for each contrast at each age. Asterisks above the bars indicate those cases in which fixation during the first test block was significantly greater than fixation during the preshift trial block.**

The divergent developmental pattern for these two contrasts cannot be explained by any difference in phonetic exposure, because neither ejectives nor clicks occur as allophones of any English consonants. For this reason (as well as those provided earlier in this report), neither can the discrepant developmental trends for these two nonnative contrasts be explained by differences in neural attrition due to differential phonetic exposure. Nor could the click versus 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 (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 our study, differential acoustic salience would be difficult to argue in any event, because these two particular nonnative contrasts are similar in their acoustic properties, as described earlier.

We suggest, instead, that the answer can be found in the development of the linguistically motivated 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, 1994a, 1994b, in press; 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 nonnative segmental contrasts. However, at least by sometime in the second half of their 1st year, infants begin to recognize certain basic characteristics of the native phonological system, which in turn begins to influence their perception of nonnative segmental contrasts. Of course, it remains to be determined exactly which characteristics of the phonology infants begin to recognize at this time—that is, what constitutes the infant's knowledge of the native phonological system at this point in development—and what permits the infants to shift from a prelinguistic to a more truly linguistic basis for perceiving speech.

These findings are, nonetheless, consistent with the notion that discrimination of a nonnative 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 nonnative consonants fit the definition of a NA contrast from the adult's perspective. Such was the case here for the Zulu click consonants. If, on the other hand, the nonnative consonants fit the definition of a SC assimilation type for adults, 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, however, it is unclear whether infants actually assimilate nonnative consonants into native phonological categories in the same way that adults do, or even whether they have fully-specified phonological categories (Werker & Pegg, 1992). It would be reasonable to suppose that the answer to these 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 adultlike levels of language competence. Additional research will be needed to address the issues of nonnative phonetic assimilation and development of phonological categories in infants.

Interestingly, the data gathered in this study with the conditioned fixation procedure revealed a hint of a language-specific effect in perception of a nonnative 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, 1986) 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 of origin for click consonants, and in those Bantu languages that borrowed the clicks from Khoisan-speaking groups over centuries of frequent interaction (Herbert, 1990). Linguists have posited that phonological elements which occur frequently across languages should be easy to perceive and produce; conversely, phonological elements that are rarely found are expected to be difficult to perceive and produce (e.g., 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 "there is much confusion" within the class of click consonants (Herbert, 1990, p. 123) when nonclick 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 target Khoisan languages (Herbert, 1990). Additional evidence suggests articulatory difficulties with the production of clicks as phonological elements. For the 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 lateral and apical clicks (Herbert, 1983).

The perceptual findings from this study are consistent with predictions based on PAM that there should be divergent developmental paths for perception of different types of nonnative consonant contrasts. This study supported the hypothesis that discrimination would show a developmental decline for a nonnative contrast which 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, which they hear as nonspeech sounds. Further research will be needed, however, to assess actual assimilation patterns in infants and to corroborate PAM's discrimination predictions for other types of assimilation patterns, including TC (two category) and CG (category goodness difference) assimilation types.

## REFERENCES

Aslin, R.N., & Pisoni, D.B. (1980). Some developmental processes in speech perception. In G. Yeni-Komshian, J.F. Kavanaugh, & C.A. Ferguson (Eds.), *Child phonology, Volume 1: Production* New York: Academic.

Best, C.T. (1984). Discovering messages in the medium: Speech and the prelinguistic infant. In H.E. Fitzgerald, B. Lester, & M. Yogman (Eds.), *Advances in pediatric psychology* (Vol. 2). New York: Plenum.

Best, C.T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Best, C.T. (1994a). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. Goodman & H.C. Nusbaum (Eds.), *The development of speech per-*

*ception: The transition from speech sounds to spoken words.* Cambridge, MA: MIT Press.

Best, C.T. (1994b). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research.* Timonium, MD: York Press.

Best, C.T. (in press). Learning to perceive the sound pattern of English. In C. Rovee-Collier & L. Lipsitt (Eds.), *Advances in infancy research* (Vol. 8). Norwood, NJ: Ablex.

Best, C.T., McRoberts, G.W., & Sithole, N.N. (1988). The phonological basis of perceptual loss for non-native contrasts: Maintenance of discrimination among Zulu clicks by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 345–360.

Best, C.T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20,* 305–330.

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.

Diehl, R., & Kluender, K. (1989). On the objects of speech perception. *Ecological Psychology, 1,* 1–45.

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). Developmental changes in speech discrimination in infants. *Journal of Speech and Hearing Research, 20,* 766–780.

Eimas, P.D. (1975). Speech perception in early infancy. In L.B. Cohen & P. Salapatek (Eds.), *Infant perception: From sensation to cognition.* New York: Academic.

Eimas, P.D., & Miller, J.L. (1980). Discrimination of the information for manner of articulation by young infants. *Infant Behavior and Development, 3,* 367–375.

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

Flege, J.E. (1989). Chinese subjects' perception of the word-final English /t/–/d/ contrast: Before and after training. *Journal of the Acoustical Society of America, 86,* 1684–1697.

Flege, J.E. (in press). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues.* Timonium, MD: York Press.

Flege, J.E., & Eefting, W. (1987). The production and perception of English stops by Spanish speakers of English. *Journal of Phonetics, 15,* 67–83.

Herbert, R.K. (1983, July). *Clicks in normal and delayed acquisition of Zulu.* Paper presented at the International Congress of Phonetic Sciences, Utrecht, The Netherlands.

Herbert, R.K. (1990). The relative markedness of click sounds: Evidence from language change, acquisition, and avoidance. *Anthropological Linguistics, 32,* 120–138.

Jakobson, R. (1958). What can typological studies contribute to historical comparative linguistics? *Proceedings of the Eighth International Congress of Linguists.* Oslo, Norway: Oslo University Press.

Jusczyk, P.W. (1992). Developing phonological categories from the speech signal. In C.A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications.* Timonium, MD: York Press.

Jusczyk, P.W. (1993). Sometimes it pays to look back before you leap ahead. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Jusczyk, P.W. (in press). Language acquisition: Speech sounds and the beginnings of phonology. In J.L. Miller & P.D. Eimas (Eds.), *Handbook of perception and cognition. Vol. 11: Speech, language, and communication.* Orlando, FL: Academic.

Jusczyk, P.W., Kemler Nelson, D.G., Hirsh-Pasek, K., Kennedy, L., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology, 24,* 252–293.

Jusczyk, P.W., & Thompson, E. (1978). Perception of a phonetic contrast in multisyllabic utterances by 2-month-old infants. *Perception and Psychophysics, 23,* 105–109.

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

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255,* 606–608.

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

Lindblom, B., Krull, D., & Stark, J. (1993). Phonetic systems and phonological development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Lindblom, B., & Maddieson, I. (1988). Phonetic universals in consonant systems. In L.M. Hyman & C.N. Li (Eds.), *Language, speech, and mind.* London & New York: Routledge.

Lisker, L., & Abramson, A.S. (1970). The voicing dimension: Some experiments on comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences.* Prague: Academia.

Lively, S.E., Logan, J.S., & Pisoni, D.B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories.

*Journal of the Acoustical Society of America, 94,* 1242–1255.

Logan, J.S., Lively, S.E., & Pisoni, D.B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89,* 874–886.

Louw, J.A. (1964). The consonantal phonemes of the lexical root in Zulu. *Afrika und Übersee, 48,* 127–152.

MacKain, K.S. (1982). Assessing 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.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C.A. (1988). A precursor of language acquisition in young infants. *Cognition, 29,* 143–178.

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 and Psychophysics, 18,* 331–340.

Pisoni, D.B., Aslin, R.N., Perey, A.J., & Hennessey, 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.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America, 89,* 2961–2977.

Polka, L. (1992). Characterizing the influence of native experience on adult speech perception. *Perception and Psychophysics, 52,* 37–52.

Polka, L. (1995). Linguistic influences in adult perception of non-native vowel contrasts. *Journal of the Acoustic Society of America, 97,* 1286–1296.

Polka, L., & Werker, J.F. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 421–435.

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

Swoboda, P.J., Kaas, J., Morse, P.A., & Leavitt, L.A. (1978). Memory factors in infant vowel discrimination in normal and at-risk infants. *Child Development, 49,* 332–339.

Swoboda, P.J., Morse, P.A., & Leavitt, L.A. (1976). Continuous vowel discrimination in normal and at-risk infants. *Child Development, 47,* 459–465.

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. (1973). Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology, 9,* 91–96.

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

Walley, A.C., Pisoni, D.B., & Aslin, R.N. (1981). The role of early experience in the development of speech perception. In R.N. Aslin, J.R. Alberts, & M.R. Petersen (Eds.), *Development of perception* (Vol. 1). New York: Academic.

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

Werker, J.F., & Lalonde, C.E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology, 24,* 672–683.

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

Werker, J.F., & Pegg, J. (1992). Infant speech perception and phonological acquisition. In C.A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications.* Timonium, MD: York Press.

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.

Williams, L. (1979). The modification of speech perception and production in second language learning. *Perception and Psychophysics, 26,* 95–104.

02 March 1994; Revised 26 July 1994 ■