# Cross-Language Speech Perception: Initial Capabilities and Developmental Change

Janet F. Werker and Chris E. Lalonde
University of British Columbia, Vancouver, Canada

This article reports three studies designed to increase our understanding of developmental changes in cross-language speech perception. In the first study, we compared adult speakers of English and Hindi on their ability to discriminate pairings from a synthetic voiced, unaspirated place-of-articulation continuum. Results indicated that English listeners discriminate two categories (ba vs. ɖa), whereas Hindi listeners discriminate three (ba vs. ɖa, and ɖa vs. Ḍa). We then used stimuli from within this continuum in the next two experiments to determine (a) if our previously reported finding (Werker & Tees, 1984a) of a reorganization between 6 and 12 months of life from "universal" to "language-specific" phonetic perception would be evident using synthetic (rather than natural) stimuli in which the physical variability within and between categories could be controlled, and (b) whether the younger infants' sensitivity to nonnative speech contrasts is best explained by reference to the phonetic relevance or the physical similarity of the stimuli. In addition to replicating the developmental reorganization, the results indicate that infant speech perception is phonetically relevant. We discuss the implications of these results.

Since the early 1970s researchers have been studying infant speech perception, partly in an attempt to determine if there is evidence for a specialized phonetic mode of processing among infants. In a now-classic study, Eimas, Siqueland, Jusczyk, and Vigorito (1971) demonstrated that English-learning infants aged 1 to 4 months show far better discrimination along a synthetic voice onset time (VOT) continuum for two stimuli that straddle the adult /ba/–/pa/ phonetic category boundary than they do for two equally acoustically distinct stimuli from within the same phonetic category. This differential discrimination at some, rather than other, points along a single synthetic continuum indicates that speech perception may be "categorical" in infants. This finding has now been replicated many times with infants using several other distinctions (for reviews, see Aslin, 1987; Kuhl, 1987). More recent research indicates that infants form categories across a variety of acoustic contexts (Miller & Eimas, 1983). For example, Kuhl and her colleagues (Kuhl, 1979, 1983) have shown that infants can categorize stimuli according to vowel color across discriminably different variations in speaker and pitch contour. Similar phonetic perceptual constancy has been shown for consonants (Hillenbrand, 1983; Kuhl, 1985).

The claim for specialized phonetic processing in the young infant has been challenged on several counts, however. It is thought to be weakened by research that shows that infrahumans also seem to show categorical perception for human speech sounds (cf. Kuhl & Padden, 1983) and by recent work showing that some nonspeech sounds may be perceived categorically by human infants (Jusczyk, Pisoni, Walley, & Murray, 1980). On the basis of these more recent results, many researchers have rejected the notion of specialized phonetic processing in young infants. They suggest instead that, although it may be "special" in the adult, speech perception can be more parsimoniously explained by general auditory mechanisms in the young infant (cf. Aslin, 1987; Jusczyk, 1985; Studdert-Kennedy, 1986a).

In spite of disagreements over the specificity of the processing mechanism, there seems to be little argument that when young infants are presented with speech and speech-like stimuli, they show enhanced discriminability at adult phonetic category boundaries. Such performance is clearly "phonetically relevant" even though it does not specify whether the infant is using a phonetic or an auditory mode of analysis. That is, an infant could show phonetically relevant perception because of discontinuities in auditory sensitivities or because of a specialized phonetic processor. Regardless of the underlying processing mechanism (and regardless of whether other animals share these sensitivities), such perceptual biases can be seen to be advantageous to the infant in the eventual task of language learning. It is for this reason that it is important to document the extent to which speech perception is phonetically relevant in the young infant.

As we review below, the hypothesis that infant speech perception is phonetically relevant is consistent with countless studies

of English-learning infants listening to English speech stimuli, but it has only been addressed in cross-language infant research using the VOT continuum. In addition, and perhaps of greater importance, little attention has been focused on the developmental origin or fate of these early abilities until very recently. The present research seeks to address these concerns.

## Cross-Language Speech Perception and Developmental Change

One way to investigate questions regarding both the origin of phonetically relevant perception and possible developmental change in early abilities is to study cross-language speech perception. This allows us to study initial infant sensitivities prior to the onset of experience and to investigate the impact of differential language exposure (for a thorough review of the effects of experience on cross-language speech perception, see Aslin & Pisoni, 1980).

To date, cross-language research has indicated that infants can discriminate native and nonnative VOT distinctions, but are less able to discriminate VOT contrasts that are not relevant in any language (Eimas, 1975; Streeter, 1976). For example, it has been shown that infants under 4 months of age can discriminate both the /ba/-/pa/ and /pʰa/-/pa/ distinctions regardless of whether they are used in the infants' language-learning environment (Lasky, Syrdal-Lasky, & Klein, 1975; Streeter, 1976) but cannot discriminate differences in VOT that are not phonetically relevant in any language (Eimas, 1975). Findings of this sort cannot provide definitive evidence of either a phonetic or auditory mode of analysis, but the match between initial perceptual biases and the phonetic status of the stimuli suggests that initial speech perception capabilities (at least for VOT) are phonetically relevant.

The evidence that infants can discriminate nearly every nonnative phonetic contrast on which they have been tested suggests a possible universality to infant perception. In some cross-language research, young infants have been shown to be able to discriminate phonetic pairs that may occur as allophones, but are not used phonemically in their language-learning environment (Aslin, Pisoni, Hennessy, & Perry, 1981; Lasky et al., 1975; Streeter, 1976). Other cross-language research has shown young infants to be able to discriminate phonetic contrasts containing at least one phone that is not even produced as an allophone in their language-learning environment (Trehub, 1976; Werker, Gilbert, Humphrey, & Tees, 1981). Although the research indicates that some phonetic contrasts may be perceptually easier for young infants than others (see Aslin et al., 1981; Best, McRoberts, & Sithole, 1988; Eilers, Gavin, & Wilson, 1979), it is clear that, if sensitive-enough procedures are used, young infants can discriminate nearly every phonetic contrast, native or nonnative, on which they have been tested.

In contrast to this high level of infant ability, research has shown that adult subjects more easily perceive those phonetic contrasts that are phonemic, that is, those that are used to differentiate meaning in their native language (Lisker & Abramson, 1970; MacKain, Best, & Strange, 1981; Trehub, 1976). It was once believed that adults had permanently lost the ability to discriminate nonnative phonetic contrasts, but more recent research has shown that adults still possess or can reacquire this

ability if they are given enough training (Jamieson & Morosan, 1986; Pisoni, Aslin, Perey, & Hennessy, 1982), are tested in a sensitive-enough procedure (Werker & Tees, 1984b), or spend enough time learning a language in which that contrast is used (MacKain, Best, & Strange, 1981). In addition, the perceptual or phonological status (or both) of the contrast in question seems to influence the ease of adult discriminability, with some nonnative distinctions being almost immediately discriminable (Best et al., 1988; Polka, 1987) and others, considerably more difficult. Nevertheless, it is safe to conclude, even with the easiest distinctions, that adults discriminate native phonetic contrasts with less difficulty than they do nonnative (for reviews, see Burnham, 1986; Strange, 1987).

This pattern of research findings has led to the hypothesis, first suggested by Eimas (1975), that infants may have a biological predisposition to discriminate the universal set of phonetic contrasts, and that there is an apparent decline or reorganization in this universal phonetic sensitivity as a function of learning a particular language. In our previous work (Werker & Tees, 1984a), we provided data that are consistent with this hypothesis and that suggest that the perceptual reorganization evident between infancy and adulthood may occur, or at least begin, within the first year of life.

In a series of experiments in our laboratory (Werker & Tees, 1983; 1984a) we tested infants, children, and adults on their ability to discriminate multiple repetitions of naturally produced syllables taken from one English (/ba/ vs. /da/) and two non-English minimal pair contrasts. The two non-English contrasts used were the Hindi voiceless, unaspirated retroflex/dental place-of-articulation (/Ta/-/ta/) and the Hindi dental voiceless-aspirated versus breathy voiced distinction (/tʰa/-/dʰa/). We used multiple naturally produced repetitions of each phoneme. This differed from traditional cross-language discrimination research (including that reviewed above), in which only one exemplar from each category is typically used. The use of multiple exemplars requires subjects to both discriminate between phoneme categories and to disregard within-category variability.[1] As a result, the task becomes a more direct test of the categorical nature of infant speech perception (see Kuhl, 1979, 1983; Miller & Eimas, 1983).

Results indicated that both English-learning infants and Hindi-speaking adults could discriminate these Hindi contrasts, but that English-speaking adults (Werker et al., 1981) and children age 4 and older (Werker & Tees, 1983) had more trouble, particularly with the perceptually more difficult place distinction. Limited training facilitated discrimination of the nonnative voicing distinction, but was totally ineffective at facilitating discrimination of the non-English retroflex/dental place distinction in this procedure (Werker et al., 1981). This set of results has been replicated using an additional non-English pla-

---

[1] As a result, this task is closer to the categorization tasks used by Kuhl (1983, 1985). It should be noted, however, that in Kuhl's work infants were required to treat "discriminably different stimuli as equivalent." We have not yet tested infants to see if they can discriminate the several exemplars from each category and, therefore, do not refer to this task as a categorization task. We do, however, have data for adults showing discrimination of this within-category variation (Werker & Tees, 1984b; Werker & Logan, 1985).

ce-of-articulation distinction, the Interior Salish (Thompson or, in its native term, Inslekepmx̂) glottalized velar versus uvular distinction (/k̓i/–/q̓i/) (Werker & Tees, 1984a).

Results from subsequent cross-sectional and longitudinal studies indicated that English-learning infants 6 to 8 months of age could discriminate both the Hindi and the Thompson place-of-articulation distinctions, but, by 10 to 12 months of age, very few of the English-learning infants could discriminate either of these distinctions. However, when Thompson and Hindi-learning infants 11 to 12 months of age were brought into the laboratory, they were still capable of discriminating these contrasts, which are, of course, phonemic in their language-learning environments (this study is described in much greater detail in Werker & Tees, 1984a).

Note, however, that in our previous research (Werker & Tees, 1984a) we tested infants on their ability to discriminate multiple repetitions of naturally produced syllables taken from non-English phonetic contrasts. Although we chose these natural exemplars so that the distribution of nonphonetic acoustic cues such as fundamental frequency, duration, and so forth overlapped between categories, as would be expected in naturally produced syllables, the distribution of phonetically relevant information was nonoverlapping. For example, *starting* frequency for the second formant was quite distinct between retroflex and dental exemplars. It varied between 1660 and 1700 Hz for the retroflex exemplars and between 1430 and 1500 Hz for the dental exemplars. This within-category similarity also applied to other acoustic cues, such as the starting frequency of the third formant and the frequency range and intensity of the burst (acoustic characteristics can be found in Werker & Tees, 1984b, or Werker & Logan, 1985).

This raises the possibility that, although the younger infants could discriminate the several stimuli according to the retroflex and dental phonetic categories, they may have done so on the basis of physical or acoustic differences independent of the phonetic status of the stimuli. This would suggest that 6-month-olds are simply more easily able to group acoustic stimuli according to any kind of physical similarity than are older infants and adults and would call into question the assertion that infant speech perception is even phonetically relevant.

This possibility is particularly problematic because the bulk of existing cross-language research has focused on VOT distinctions. Indeed, all of the cross-language research showing that infants can *only* discriminate phonetically relevant contrasts has been confined to studies of single-exemplar discriminations along the VOT continuum. As such, it is important to ascertain whether the phonetic specificity of infant perception generalizes to procedures that use multiple exemplars and other kinds of contrasts.

The present research sought to address this problem. To separate the isomorphism between physical similarity and phonetic relevance present in our previous work (Werker & Tees, 1984a), we used synthetic stimuli. The first experiment was designed to test the validity of these synthetic stimuli. Experiments 2 and 3 were designed to determine whether the younger infants' ability to discriminate nonnative contrasts could be attributed to sensitivity to physical similarity regardless of phonetic status, as suggested above, or whether the phonetic relevance of the stimuli is a more potent factor in explaining infant

sensitivity. This was assessed in Experiments 2 and 3 by the inclusion of three kinds of speech contrasts: an English contrast, a non-English contrast, and a contrast that is not used in any language. Second, Experiments 2 and 3 were designed to determine if our previously reported reorganization between 6 and 12 months would be evident using a new, but related, speech contrast.

## Experiment 1

This experiment was designed to determine if it is possible to synthesize a single place-of-articulation continuum that English listeners reliably divide into two categories, bilabial and alveolar, and that Hindi adults reliably divide into three categories, bilabial, dental, and retroflex. Voiced stimuli were used as a means of further testing the generalizability of our previous cross-language research (Werker & Tees, 1984a) in which we used natural, voiceless place distinctions.

Previous research has shown cross-language differences in the perception of synthetic voicing (Flege, 1984; Flege & Hillenbrand, 1986; Lisker & Abramson, 1970; Williams, 1980) and manner (MacKain et al., 1981) continua, but there has been little research with respect to cross-language differences in place. The one notable exception was a recent study by Lisker (1985). In this study, Lisker synthesized two 3-formant, place-of-articulation continua, one for voiced and one for voiceless unaspirated Hindi stop consonants. Theoretically, the stimuli should span the bilabial, dental, and retroflex phonetic categories for Hindi listeners. Stimuli from this continuum were then presented to three native Hindi speakers. The Hindi listeners reliably labeled the bilabial and dental stimuli, but were fairly unreliable in their labeling of retroflex consonants. Lisker reported much greater reliability for the voiced than for the voiceless unaspirated series, providing an additional reason for using a voiced rather than voiceless series in the present research.

The present study can be seen as an extension of this earlier work by Lisker (1985), one that uses a newly synthesized continuum. The purpose of this experiment was to determine if there are cross-language differences in discrimination along this continuum. Prior to testing for discrimination, we asked both English and Hindi listeners to identify randomly presented stimuli from the voiced place-of-articulation continuum to ensure that members of each language group could reliably label these stimuli according to the phonemic categories used in their own native language. We then tested English and Hindi subjects on their ability to discriminate all the two-step pairings using an ABX testing procedure to determine if labeling performance would predict discrimination.

### Method

*Subjects.* Ten Hindi-speaking and 10 English-speaking adults served as subjects in this experiment. The 5 male and 5 female English subjects ranged in age from 18 to 45 years. The 7 male and 3 female Hindi subjects ranged in age from 18 to 29 years. All 10 of the English-speaking subjects lived in or around Halifax, Nova Scotia. Eight of the 10 English subjects had some familarity with French, but none spoke it fluently; the other 2 subjects spoke only English.

Six of the native Hindi-speaking subjects lived in the Halifax area, and 4 lived in Burnaby, British Columbia. All subjects reported that
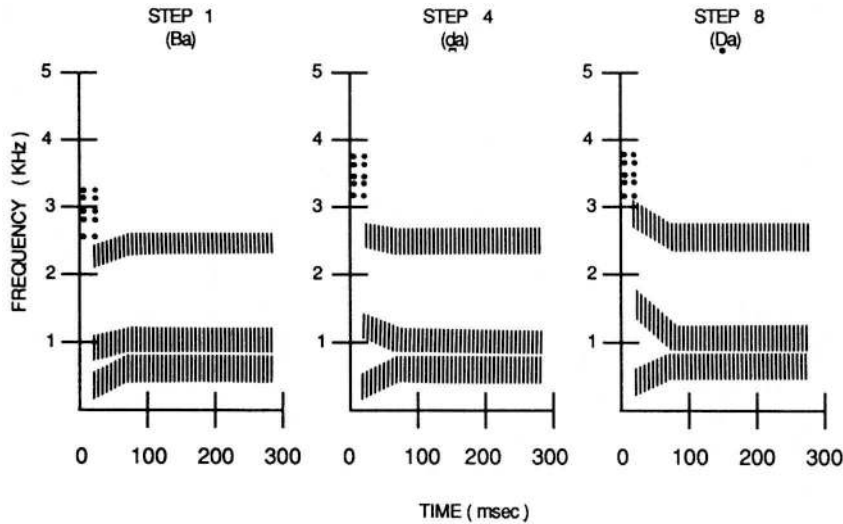
*Figure 1.* Sample pseudospectrograms of the synthesizer input parameters.

Hindi was their first language and all also spoke English. Eight also reported that they spoke Punjabi fluently, 1 also spoke Gujarti, 1 spoke Urdu, and 2 spoke Bengali. All native Hindi speakers had lived in India for at least the first 17 years of their life, and 8 had been in North America for less than 3 years. The other 2 Hindi speakers had been in North America for 11 and 12 years, but each had returned to India for periodic visits during that time. All native Hindi speakers reported that they still spoke Hindi at least 20% of the time.

*Stimuli.* The synthetic voiced place-of-articulation continuum was constructed by considering the work of Lisker (1985), Stevens and Blumstein (1975), and Ladefoged and Bhaskararao (1983) on retroflex and dental-stop consonants as well as an acoustic analysis of the naturally produced initial-position voiceless/unaspirated dental and retroflex syllables used in our previous work (see Werker & Tees, 1984b). The vowel /a/ was selected because it is widely distributed after bilabial, dental, alveolar, and retroflex stops.

Five formant stimuli were constructed with the use of the Mattingly synthesizer on the VAX 11/780 at Haskins Laboratories, New Haven, Connecticut. An eight-step continuum was constructed by varying the starting frequency of F2 and F3 (second and third formants).

Following the work of Stevens and Blumstein (1975), the burst was 10 ms in duration and directly preceded the onset of voicing.[2] The frequency range for the burst was lower for the first two steps on the continuum than for the remaining six steps. This was necessary to produce stimuli that would be perceived as coherent syllables (otherwise the burst would perceptually separate from the syllable). Pilot testing indicated that, beyond the first two steps, it made no difference to the coherence of the percept whether the frequency range of the burst continued to increase along with the continuing increase in the starting frequency of F3. Thus, to keep the steps in the continuum equivalent on as many cues as possible, the frequency range of the burst was held constant throughout the remaining six steps.

Each stimulus was 275 ms in duration. Fundamental frequency was steady at 100 Hz for the first 100 ms, then gradually rose to 120 Hz during the remaining 175 ms. F1 rose from 250 to 500 Hz in a 50-ms transition, and F4 and F5 were constant throughout the continuum at 3500 and 4000 Hz, respectively. The starting frequency of F2 varied in eight equal 100-Hz steps from 900 to 1600 Hz, and the starting frequency of F3 varied in 96-Hz steps from 2240 to 2912 Hz. Transitions for F2 and F3 were 50 ms. The steady state frequency was 1090 Hz for F2 and 2440 Hz for F3. Sample pseudospectrograms of the synthesizer input parameters are shown in Figure 1.[3]

*Procedure and apparatus.* Tape-recorded stimuli were played on a Teac 3340S reel-to-reel tape recorder to a single driver speaker. Subjects were tested individually in an IAC sound-attenuated room.

In the Identification task there were eight repetitions of each stimulus presented in random order, for a total of 64 trials. The intertrial interval (ITI) was 3500 ms. Subjects were told that they would be hearing a list of short syllables. They were instructed to listen to each syllable carefully and circle the appropriate answer on a prepared answer sheet. Because the purpose of this first task was simply to see if Hindi and English speakers could consistently label these stimuli according to native-language phonemic categories, Hindi subjects were given three possible category labels corresponding to the Davanagari characters for voiced unaspirated bilabial ( ब ), dental ( द ) and retroflex ( ड ) stops in the presence of a long /a/ vowel. In accord with their phonemic categories, English subjects were given only two possible category labels, (ba) and (da).

We used an ABX discrimination task to assess possible cross-language differences in discrimination. In the ABX discrimination task, subjects are presented with triads of syllables and are required to indi-

---

[2] In earlier pilot studies, we used the precise stimulus characteristics described by Stevens and Blumstein (1975) for retroflex and dental stimuli. However, the Hindi informants reported that all these stimuli sounded like retroflex rather than dental stops. In subsequent synthesis attempts, stimuli were thus modeled more closely on the naturally produced retroflex and dental stops used in our previous work (Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Logan, 1985; Werker & Tees, 1984a).

[3] Both extensive pilot testing using a free-labeling task and an acoustic analysis of our natural tokens indicated that brief (50 ms) transitions and a fairly intense burst were needed for Hindi listeners to perceive three places along the continuum. Without a burst, or with a weak burst, Hindi listeners divided the continuum into only two categories—bilabial and dental. With a burst, but with longer (70 ms) F2 or F3 transitions, Hindi listeners still perceived three categories, but with much less certainty. Certainty also diminished for English listeners at the extreme *ba* end of the continuum when either the burst was absent or the transitions lengthened (some subjects reported a *wa* sound). Lengthening F3 also reduced certainty at the *da* end of the continuum for English listeners. Inclusion of a burst changed the boundary location by one step for both Hindi and English listeners.
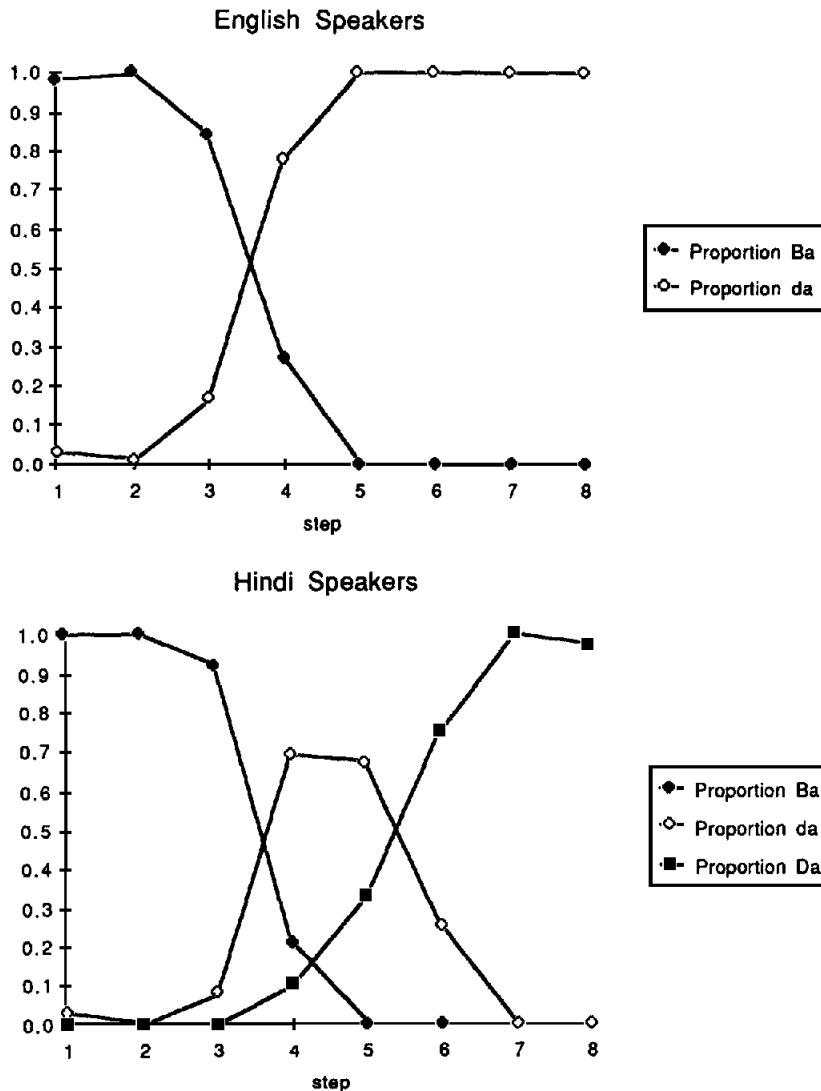
English Speakers



Hindi Speakers



*Figure 2.* Average group identification functions.

cate whether the third syllable sounded more like the first or the second syllable by circling *A* or *B* on a prepared answer sheet. The inter-stimulus interval (ISIs) within trials were 1000 ms. The ITIs were 3500 ms. Each possible two-step pairing was presented 16 times: 4 times in each order with the correct answer balanced for each order, for a total of 96 randomly ordered presentations.

## Results

*Identification task.* Performance in the Identification (labeling) test revealed that Hindi-speaking adults can reliably and consistently divide this continuum into the three categories used in their native language, and that English-speaking adults can reliably and consistently divide it into two. The average group Identification functions are shown in Figure 2. It can be seen that both Hindi and English speakers label the first three stimuli from the continuum /ba/. English listeners consistently labeled Stimuli 5 to 8 /da/. The majority labeled Stimulus 4 as

/da/, but some labeled it /ba/. Hindi listeners consistently labeled Stimuli 4 and 5 as dentals /d̪a/, and Stimuli 7 and 8 as retroflex /Da/, but showed considerable variability on Stimulus 6, labeling it both dental and retroflex.[4]

*ABX discrimination task.* The critical research question was whether there are cross-language differences in discrimination in accord with the category labels that are available to Hindi and English speakers. Predicted discrimination scores were computed for each subject using the formula outlined in Mac-

---

[4] It should be noted that similar functions were obtained in the free-labeling task used in the pilot studies. In that task, final stimuli were labeled as either /ba/ or /da/ over 90% of the time by English listeners and were labeled as either bilabial, dental, or retroflex by Hindi listeners. In the free-labeling task, however, Hindi listeners on occasion would label the stimuli as voiced-aspirated rather than voiced-unaspirated stops.

Table 1
*Predicted and Obtained ABX Discrimination Scores*

| Language | Pairing | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| | 1–3 | 2–4 | 3–5 | 4–6 | 5–7 | 6–8 |
| English | | | | | | |
| Predicted | .53 | .80 | .88 | .58 | .50 | .50 |
| Obtained | .53 | .77 | .86 | .61 | .50 | .48 |
| Hindi | | | | | | |
| Predicted | .53 | .79 | .86 | .69 | .73 | .53 |
| Obtained | .47 | .79 | .88 | .66 | .69 | .51 |

millan, Kaplan, and Creelman (1977). Average predicted and obtained discrimination scores for each pairing by group were then computed. These values are shown in Table 1. As can be seen, there is an excellent fit between predicted and obtained values for both English and Hindi listeners. There is only one discrimination boundary for English listeners, and there are two for Hindi listeners.

The data were then probed in several ways. To test for the goodness of fit between predicted and obtained discrimination functions, we ran a series of $t$ tests comparing obtained and predicted discrimination scores for each pairing for each group. As would be expected from an examination of Table 1, there were no significant differences, suggesting a good fit between obtained and predicted discrimination scores.

Second, we ran a series of $t$ tests comparing performance on each pairing with a hypothetical chance performance of .5. On the basis of the labeling data, it was predicted that English adults would perform significantly better than chance only on Pairings 2 versus 4 and 3 versus 5, because these both cross the English bilabial/alveolar boundary. All the other pairings are within category for English listeners. It was predicted that Hindi listeners would perform significantly better than chance on Pairings 2 versus 4, 3 versus 5, 4 versus 6, 5 versus 7, with Pairings 4 versus 6 and 5 versus 7 corresponding to the dental/retroflex boundary. All the predicted differences were significant at the .01 level except the Hindi Pairing 4 versus 6, which was significant at the .05 level. No other pairings were significantly different from chance.

In a final analysis, the performance of Hindi and English listeners was compared across all pairings. Hindi adults did not perform significantly differently than English adults on any pairing except 5 versus 7 (the pairing that most clearly crosses the Hindi dental/retroflex boundary and is within the alveolar category for English speakers). The Hindi adults performed better than the English adults on this pairing, $t(18) = 4.96$, $p < .001$.

The results from the ABX discrimination task indicated that there are cross-language differences in discrimination performance along this continuum, and that these differences can be predicted on the basis of the category labels that are available in each language. Specifically, English listeners showed evidence of two categories in their Identification functions and one boundary in their ABX discrimination functions, whereas Hindi listeners showed evidence of three categories and two discrimination boundaries.

## Experiment 2

The second experiment was designed to determine (a) whether our original finding (Werker & Tees, 1984a) of a reorganization from universal to language-specific perception in the first year of life extends to new subjects and a new (but related) non-English contrast, and (b) whether perception is constrained to phonetically relevant categories in the young infant, or whether infants can discriminate between any two sets of adjacent stimuli, regardless of whether the sets represent different adult phonetic categories.

In our previous research (Werker & Tees, 1984a), we showed that young infants (but not older infants and adults) can discriminate between native and nonnative phonetic categories. As outlined in the introduction, we tested infants using multiple repetitions of naturally produced exemplars from the Hindi retroflex and dental phonetic categories. One problem apparent in this earlier research was that the dimensions of phonetic relevance and physical similarity were confounded. That is, with the natural repetitions, the physical variability within category was much less than that between categories. As such, it is possible that the young infants were actually discriminating the stimuli on the basis of physical differences irrespective of phonetic status.

The synthetic place-of-articulation continuum that was developed for Experiment 1 provided a set of stimuli that enabled us to separate the dimensions of physical similarity and phonetic relevance. In this endeavor, the synthetic continuum was half-stepped, yielding a 16-step continuum. To test for the phonetic relevance of speech perception, we required subjects to discriminate three adjacent stimuli from either side of three locations along the continuum. These locations were labeled *Common, Hindi-only,* and *Neither.*

The Common location corresponds to the boundary between the bilabial and medial (dental or alveolar) stops that is common to both English and Hindi speakers. The Hindi-only location corresponds to the boundary between the Hindi dental and retroflex categories, and the Neither location is at an arbitrarily chosen location near the *Da* end of the continuum, which is not known to distinguish phonetic categories in any of the languages of the world. The stimuli selected represent equal step intervals across three locations, resulting in a set of stimuli with equivalent within- and between-category variability.

We tested four groups of subjects on their ability to discriminate at each of the locations: Common, Hindi-only, and Neither. This allowed us to test specific predictions about the initial phonetic relevance and subsequent developmental fate of infant speech-perception capabilities. If perception is phonetically relevant in young infants and is constrained to phonetically relevant stimuli, it would be predicted that young infants (and Hindi adults) should be able to discriminate stimuli at the Common and the Hindi-only, but not at the Neither, locations. Conversely, if speech perception in the young infant is organized on the basis of physical similarity irrespective of phonetic relevance, young infants should be able to discriminate at the Neither location as well. With regard to developmental predictions, if there is a reorganization within the first year of life, older infants and English-speaking adults should only be able to dis-

criminate stimuli according to the Common, but not the Hindi-only or Neither, location.

## Method

*Subjects.* Group 1 included 18 English-speaking adults, 3 tested in each of six possible orders. These subjects were all first-year psychology students at the University of British Columbia who received course credit for their participation.

Group 2 included 5 Hindi-speaking adults. These subjects were all students or postdoctoral fellows at the University of British Columbia. All Hindi-speaking subjects filled out a language-experience questionaire. Four of the 5 subjects grew up in India, and 1, in Malaysia. Although all spoke Hindi fluently and had done so since at least age 7, 3 reported that their first language was Punjabi, 1 that it was Telugu, and only 1 that it was Hindi. One of the subjects also spoke Urdu, 1 spoke Kannada, and 1 spoke several other languages besides Punjabi and Hindi (Malay, Urdu, Tamil, and Cantonese). All Hindi-speaking subjects had been living permanently in North America for less than 4 years at the time of testing.

Group 3 included 8 infants aged 6 to 8 months (*M* age = 7 months, 4 days; range = 6 months, 23 days to 7 months, 29 days). We collected data from 12 infants but eliminated the data for 4 of them because of failure to reach initial conditioning criterion, inability to return to the laboratory for 3 days of testing, or equipment failure.

Group 4 included 8 infants aged 11 to 13 months (*M* age = 12 months, 9 days; range = 11 months, 5 days to 13 months, 24 days). We collected data from 13 infants and eliminated data for 5.

*Stimuli.* Stimuli were selected from the synthetic place-of-articulation continuum used in Experiment 1. The original 8-step continuum was further subdivided into 16 equal steps, as illustrated in Figure 3. (It should be noted that stimulus 2a had the burst characteristics of Stimuli 3 versus 8.) Three adjacent stimuli were then selected from either side of three full 1-step locations along the continuum.[5] As described in the introduction to this experiment, these correspond to the Common, Hindi-only, and Neither (adult) boundary locations. The stimuli used in each pairing are shown in Table 2.

*Procedure.* We tested infants using a variation of the head-turn procedure (for a detailed description of this procedure, see Kuhl, 1987). In our use of this procedure the infant was presented with a stream of repeating stimuli from one speech sound category (the background category) and was conditioned to turn his or her head away from an experimental assistant and toward a visual reinforcer when there was a change in the speech sound category (to the contrasting category). Correct head turns (hits) were visually reinforced by the illumination and activation of toy animals inside a smoked Plexiglas box, and incorrect head turns (false positives) were not reinforced. In addition, social reinforcement was provided by smiles, hand clapping, and the pronouncement "good (child's name)."

The three stimuli from each category were set up in random order on one of two audio tracks, with the track that served as the background counterbalanced across subjects. For example, in the Common pairing, the bilabials (2a, 3, and 3a) were set up in random order on one track, and the dental/alveolars (4a, 5, and 5a), in random order on the other
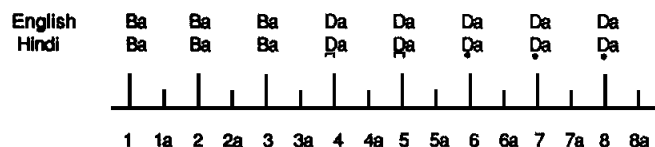
### Table 2
*Stimuli Used in Experiments 2 and 3*

| Pairing | Track 1 | Track 2 |
|---|---|---|
| Common | [2a 3 and] 3a | 4a [5 and 5a] |
| Hindi-only | [4a 5 and] 5a | 6a [7 and 7a] |
| Neither | [5a 6 and] 6a | 7a [8 and 8a] |

*Note.* Figures within brackets indicate stimuli used in Experiment 3.

track. Stimuli were presented using a Teac X2000 M reel-to-reel tape recorder. Individual stimuli were 275 ms in duration, separated by a silent interstimulus interval of 1500 ms.

In the experimental set-up for the head-turn procedure, the infant sat on his or her parent's lap in a sound attenuated room facing an experimental assistant (E2) across a table. The speaker and the visual reinforcer were located at a 100° angle, 50 cm to the right of the parent and infant. Both the parent and E2 wore headphones that delivered music so that they could not influence the infant's behavior. E2 kept the infant looking in his or her direction by manipulating small toys. Another experimenter (E1) sat outside the chamber observing the infant through a one-way glass.

E1's first task was to indicate (by pressing one of two foot pedals) when the infant was in a state of readiness—that is, watching E2, not staring at the loud speaker or reinforcer, and not fussing or babbling. When E1 signaled this state of readiness, the computer (an Apple II+) selected either a control or an experimental trial. Because trial initiation thus depended on the infant's state, the number of exemplars presented between trials varied from a minimum of 6 to over 30, preventing the infant from developing a response rhythm on the basis of timing. During control trials the infant was presented with more repetitions from the background category, and during experimental trials, with exemplars from the contrasting category. A light outside the chamber would come on to signal E1 that it was an observation interval, but would not indicate whether it was a control or an experimental trial. Thus, E1 was blind as to whether a control or an experimental trial was occurring. E1's job was to push the other foot pedal if or when he or she detected a head turn during an observation interval.

During the conditioning phase of this procedure, all trials were change trials (that is, stimuli change from Category 1/Track 1 to Category 2/Track 2). During this phase, E1 initially activated the reinforcer immediately following a category change, with a gradually increasing interval (up to 4 s) between the sound change and activation. When the infant made three correct head turns in a row prior to activation, conditioning ended and the testing phase of the experiment began.

In the testing phase, approximately half of the trials were control (no change) trials and half were experimental (change) trials (with the constraint that no more than three control trials occur in a row). Activation of the reinforcer was contingent on the infant making a correct head turn within 4 s of the onset of a change trial. If the infant failed to respond to three consecutive change trials, conditioning was reinstated for five trials or three hits (whichever came first) before testing was continued. Only those responses that occurred during testing were used in calculating each subject's performance.

We tested infants on three separate days, always within a week. Order was counterbalanced across adults, but all infants were tested in the same order: Common infants on Day 1, Hindi infants on Day 2, and



English  Ba   Ba   Ba   Da   Da   Da   Da   Da
Hindi    Ba   Ba   Ba   Ḍa   Ḍa   Ḍa   Ḍa   Ḍa

         1  1a  2  2a  3  3a  4  4a  5  5a  6  6a  7  7a  8  8a

*Figure 3.* Subdivisions of the original 8-step synthetic continuum into 16 equal steps.

---

[5] The use of more than one stimulus in each category makes this task comparable with our previous research in which multiple natural exemplars were used from each phonetic category. However, in the present research the physical variability was precisely controlled.

Neither infants on Day 3. We recognize that this aspect of the design was not ideal. There were, however, several factors that necessitated a constant order of testing with the infant subjects. First, order was kept constant primarily because this is an operant *conditioning* procedure. One of the principles of operant conditioning is that without reinforcement conditioning will be extinguished. When infants cannot discriminate a contrast, they cannot be reinforced and subsequently often refuse to perform on other contrasts in the head-turn procedure. As such, in a within-groups design it is desirable to test infant subjects on the easiest contrasts first. The decision to keep order constant across infants was also made on the basis of an analysis of the adult data, which indicated that no order effects were present. Finally, we (and others, e.g., Kuhl, 1983) have shown that infants can be tested in this procedure for up to 7 days without a decrement in performance. The partial retest given on Days 2 and 3 (described below) provided additional evidence that the infants could still perform in the procedure on subsequent days. Although it would have been ideal to have asked the parents to return for a fourth day to the laboratory in order to retest the infant's ability to discriminate the Common contrast, in too many cases this was not possible (particularly because our subjects participated on a volunteer basis). For these reasons it was decided to proceed with an order of presentation designed to elicit optimal performance from our infant subjects.

We tested infants for approximately 30 testing trials ($M = 27.1$). Any infant who failed to meet the initial conditioning criterion (i.e., three in a row correct) on the Common pairing was excluded from the study as not being able (or interested) to perform the head-turn task, because much previous work has shown that infants of this age and younger (Eimas, 1974; Mehler, 1985) can discriminate /ba/ from /da/. On Days 2 and 3, infants were initially retested on Common to make sure they could (and would) still perform in the procedure. It should be noted that no infant failed this retest measure. On reaching three consecutive correct hits on Common (usually within five to six trials), the infants were conditioned and tested on either the Hindi-only or Neither pairings.

We tested adults in a variation of this go/no-go procedure in which they were instructed to monitor the continuous string of stimuli and to press a button whenever they detected a change. In the adult procedure a light was illuminated to indicate correct button-presses only. We tested adults in a single, 60- to 90-min session on all three pairings. They were given three initial training trials on each pairing and, subsequently, were tested for approximately 30 test trials ($M = 28.8$). The order of presentation of the three different pairings was counterbalanced to determine if there were order effects. As reported above, there was no evidence of an order effect.

## Results

The proportion of correct responses was calculated for each subject by dividing the number of hits and correct rejections by the total number of trials. The proportion of correct responses to both experimental and control trials was compared across the four groups in a split-plot design. The between-groups factor was the age/language experience group, and the repeated within-groups factor was the pairing. The proportion of correct responses by group is shown in Figure 4.

The overall analysis of variance (ANOVA) indicated that there were main effects for both the between-group factor (Group), $F(3, 35) = 17.539, p < .001$, and the within-group factor (Pairing), $F(2, 6) = 48.931, p < .001$, as well as a significant Group × Pairing interaction, $F(6, 70) = 3.178, p = .008$. Subsequent one-way repeated measures ANOVAs comparing the three pairings within each age group showed there to be significant effects for each group: for the infants aged 6 to 8 months, $F(2, 14) =$

24.818, $p < .001$; for the infants aged 11 to 13 months, $F(2, 14) = 10.880, p = .001$; for the English adults, $F(2, 34) = 27.789, p < .001$; and for the Hindi adults, $F(2, 8) = 10.500, p = .006$. Post hoc Newman–Keuls comparisons (Kirk, 1968) showed this to be accounted for by Hindi and 6- to 8-month-old subjects having achieved a greater proportion of correct responses to the Common and Hindi-only pairings than to the Neither pairing, whereas older infants and English adults achieved significantly more correct responses to the Common than they did to either the Hindi-only or the Neither pairing (with no significant difference in performance on the two latter contrasts).

In addition, we ran a series of $t$ tests in which the proportion of correct responses per group was compared with a chance value of .5 for each pairing. The results were consistent with our predictions for all subjects. For the English adults, performance was significantly above the chance value on the Common pairing only, $t(17) = .025 = 7.640, p < .001$. For Hindi adults, performance exceeded chance on both the Common, $t(4) = 5.231, p = .007$, and Hindi-only, $t(4) = 12.830, p < .001$, pairings. For the 6- to 8-month-olds, performance exceeded chance on the Common, $t(7) = 9.261, p < .001$, and the Hindi-only, $t(7) = 5.661, p = .001$, pairings. For the 11- to 13-month-olds, performance significantly exceeded chance only on the Common pairing, $t(7) = 3.601, p = .008$. No group of subjects exceeded chance on the Neither pairing.

## Discussion

The results of this experiment replicate and extend our previous cross-language work using synthetic stimuli and a new (but related) set of phonetic contrasts. The results indicate that 6- to 8-month-old infants (like Hindi adults) can only discriminate multiple stimuli at the Common and the Hindi-only, but not at the Neither, locations. That these infants could discriminate the Hindi-only, but not the Neither, pairing replicates and extends previous studies using a VOT continuum (Aslin et al., 1981; Eimas, 1975; Streeter, 1976).

These results replicate our earlier finding of a reorganization within the first year of life from phonetic to phonemic perception. The younger English-speaking infants, like the Hindi-speaking adults, were able to discriminate both the Hindi-only and Common pairings. However, the infants aged 11 to 13 months, like the English-speaking adults, were significantly less able to discriminate the Hindi-only pairing in this procedure and only showed clear discrimination of the phonetic contrast (Common) that carries phonemic significance in their language-learning environment.

## Experiment 3

This experiment was designed to determine whether infant or adult discrimination (or both) of the Neither and Hindi-only pairings would improve if the categories were made more distinct. To this end, we made two modifications to the stimulus sets. First, the number of exemplars within each category was reduced to two stimuli only a half step apart, thereby increasing the similarity of exemplars within each category. Second, the distance between categories was increased to two full steps. This
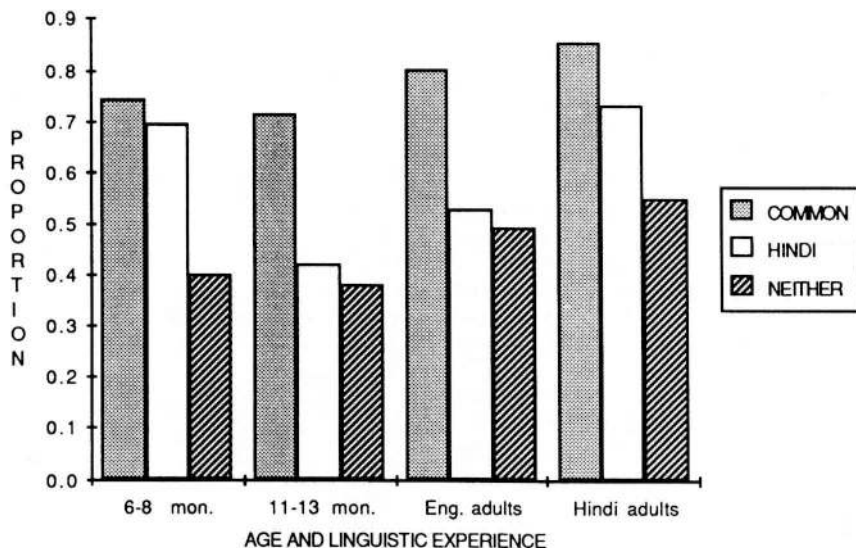
*Figure 4.* The proportion of correct responses by group for each of the one-step pairings.

resulted in a set of stimuli with reduced within-category variability (fewer exemplars) and increased between-category variability (two steps between categories)—that is, a set of stimuli biased in favor of discrimination on the basis of physical similarity.

## Method

*Subjects.* There were four groups of subjects as in Experiment 2. Group 1 included 30 English-speaking adults, 5 tested in each of six possible orders of pairings. These subjects were all psychology students at Simon Fraser University, Burnaby, British Columbia. Three subjects were eliminated from this group because of equipment failure.

Group 2 included 5 native Hindi-speaking adults. These subjects were all graduate students or postdoctoral fellows at Simon Fraser University. All 5 subjects grew up in India and reported Hindi as their first language. Four of the 5 subjects also spoke Punjabi fluently. All Hindi subjects had lived in India until at least 20 years of age and had been living permanently in North America for less than 5 years at the time of testing.

Group 3 included 8 infants aged 6 to 8 months ($M$ age = 7 months, 7 days; range = 6 months, 2 days to 7 months, 25 days). We collected data on a total of 12 infants, and eliminated 4 for the reasons outlined in Experiment 2. Group 4 included 8 infants aged 11 to 13 months ($M$ age = 11 months, 27 days; range = 10 months, 27 days to 13 months, 5 days). We collected data on 13 infants and eliminated 5 from this group.

*Stimuli.* The stimuli were once again taken from the synthesized place-of-articulation continuum. Two exemplars from either side of the Common, Hindi-only, and the Neither boundaries were selected. The cross-boundary spread was increased to two full steps. Stimulus selections are indicated in Table 2. Only the stimuli contained within the brackets were used in this experiment.

*Procedure.* The procedure and apparatus were identical to those used in Experiment 2, except for the use of a Teac 3340S reel-to-reel tape recorder.

## Results

Although the absolute numbers are different, the pattern of results is virtually identical to that obtained in Experiment 2.

The proportion of correct responses is shown in Figure 5. The proportion of correct responses was analyzed across the four groups in a split-plot design. The between factor was the age/language experience group, and the repeated within-groups factor was the pairing. The overall ANOVA indicated that there were main effects both for groups, $F(3, 44) = 4.946$, $p = .005$, and pairings, $F(2, 6) = 35.172$, $p < .001$, as well as a significant Group × Pairings interaction, $F(6, 88) = 3.187$, $p = .008$. Subsequent one-way repeated measures ANOVAS were significant for each group as follows: for 6 to 8 months, $F(2, 14) = 18.455$, $p < .001$; for 11 to 13 months, $F(2, 14) = 26.375$, $p < .001$; for English adults, $F(2, 52) = 28.033$, $p < .001$; and for Hindi adults, $F(2, 8) = 15.000$, $p = .002$. Post hoc Newman–Keuls tests were as predicted. The overall effect for the infants aged 6 to 8 months and the Hindi adults was accounted for by significantly better performance on the Common and the Hindi-only pairings than on the Neither pairings, with no differences between Common and Hindi-only. As predicted, the older infants and the English adults performed significantly better on the Common than on either the Hindi-only or the Neither contrasts, with no further significant differences.

We ran a series of $t$ tests to compare subjects' performance with a chance value of .5. The results were as predicted for the adult subjects. The performance of the English adults significantly exceeded chance only on the Common pairing, $t(26) = .025 = 9.706$, $p < .001$. The Hindi adults exceeded chance on the Common, $t(4) = 8.313$, $p = .002$, and the Hindi-only, $t(4) = 3.635$, $p = .023$, pairings. For 6- to 8-month-olds the results were as follows: for Common, $t(7) = 5.356$, $p = .001$, and for Hindi-only, $t(7) = 5.962$, $p < .001$. For 11- to 13-month-olds the results were: for Common, $t(7) = 5.325$, $p = .001$, and for Neither, $t(7) = -3.493$, $p = .011$. With the exception of this last result (indicating performance significantly below chance on the Neither pairing by 11- to 13-month-olds), all results were as predicted.

In a final set of analyses the performance of subjects in Experiment 2 was compared with that in Experiment 3. In spite of
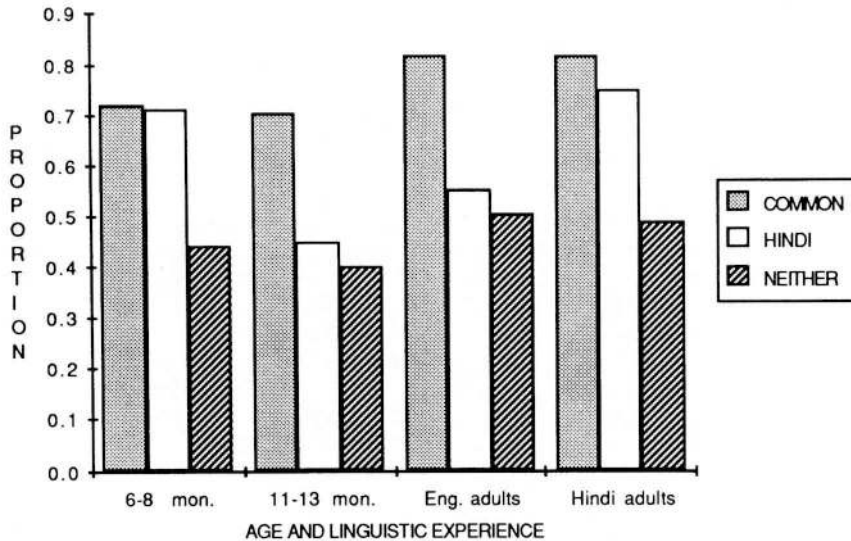
*Figure 5.* The proportion of correct responses by group for each of the two-step pairings.

apparent differences in the absolute performance of subjects between Experiments 2 and 3, no significant differences emerged for any group on any pairing between the one-step and two-step presentations.

*Discussion*

The results from Experiment 3 were consistent with those obtained in Experiment 2. That the 6- to 8-month-old infants did not discriminate the Neither pairing in Experiment 3, despite the increased physical difference between categories, strengthens the claim that infant speech perception is phonetically relevant. Furthermore, that 11- to 13-month-olds were able to discriminate the Common pairings but were not able to discriminate the Hindi-only distinction even when the between-category variability was greater than that within categories is noteworthy. This suggests that the developmental reorganization cannot be explained by a general decline in sensitivity to physically similar groupings.

The differences between Experiments 2 and 3 are not significant. This is particularly important with respect to the Neither pairings for the younger infants and the Hindi-only pairings for the older infants. Given the greater physical separation between categories, an explanation of speech perception on the basis of physical similarity would predict higher levels of discrimination on all pairings in Experiment 3. This was not the case. This is important given the considerable physical variability in the realization of an individual phoneme across speakers and across different contexts (Liberman, 1982; Liberman, Cooper, Shankweiler, Studdert-Kennedy, 1967). Of particular relevance, the within-category variability in these experiments is based on the same physical dimensions (starting frequency of F2 and F3) that signal phonetic category changes at some points along the continuum. The selective between-category discrimination shown here is consistent with a phonetically relevant interpretation of infant speech perception capabilities.

Summary and Conclusions

In summary, we ran three experiments in an attempt to replicate, clarify, and extend our previous research (Werker & Tees, 1984a). To eliminate the unequal physical variability within and between nonnative phonetic categories present in the natural stimuli used in our previous work, we constructed a synthetic place-of-articulation continuum. Experiment 1 showed that Hindi- and English-speaking adults discriminate exemplars from this continuum in accord with the phonemic categories of their own language. In both Experiments 2 and 3, infants aged 6 to 8 months were able to discriminate the Common and Hindi-only contrasts, whereas English-learning infants aged 11 to 13 months could only discriminate the Common contrast. This replicates our previous research, providing a further demonstration of the reorganization in speech perception in the first year of life. It is noteworthy that none of our infant subjects could discriminate the Neither contrast even in Experiment 3, where the absolute physical variability was greater between categories than it was within. This provides further cross-language verification that infants are sensitive to the universal set of phonetic distinctions and evidences another instance (here with regard to place of articulation) of a failure to discriminate an arbitrary contrast that is not phonetically relevant. This replicates and extends previous infant studies showing differential sensitivity to phonetically relevant contrasts along a VOT continuum. This mapping between biologically given sensitivities and phonetic categories allows the young infant to segment the incoming speech stream into discrete perceptual entities and enables the infant to divide the ongoing and overlapping stream of speech into the units that will be required in the important task of beginning to learn a language.

It should be noted once again that, although these results are consistent with the hypothesis that initial infant speech perception is phonetically relevant, they do not address the question of whether phonetically relevant sensitivities result from specifically linguistic or generalized auditory processing mecha-

nisms (cf. Aslin, 1987; Kuhl, 1987; Kuhl & Miller, 1975; Studdert-Kennedy, 1986b). It is possible that the phonetically relevant sensitivities stem from limits in the resolving power of the auditory system, so that some physical differences are simply more discriminable than others, and that human languages are organized around these auditory sensitivities (Aslin, 1987; Kuhl & Miller, 1975; Pisoni, 1979). On the basis of these data, it is also possible that the close relationship between initial (phonetically relevant) perception and the subsequent language-specific phonetic categories stems from a specialized speech-perception module (Fodor, 1983; Liberman & Mattingly, 1985; Mattingly & Liberman, in press; Whalen & Liberman, 1987).

One way to help clarify this controversy would be to test whether infants have the auditory capabilities to discriminate the several exemplars from within a single phonetic category when tested in a more sensitive procedure. If it can be shown that infants have such auditory capabilities, yet treat the within-category exemplars as equivalent, evidence in support of phonetic specificity would be strengthened. In addition, evidence for within-category discriminability would lead us toward the interpretation that this is a categorization rather than a discrimination task (Bornstein, 1987; Kuhl, 1987; Younger & Cohen, 1985). It would also be informative to determine whether adults and infants can form artificial categories that mix exemplars from two adjacent adult phonemic categories, using, for example, our Stimuli 3, 3a, and 4 as opposed to Stimuli 5, 5a, and 6. It is clear that this stimulus set would generate different predictions depending on the mechanism postulated to account for speech perception.

These data clearly indicate that by 1 year of age the perceptual capabilities of the young infant do correspond to linguistically significant categories. The developmental change between 6 and 12 months shows that the perceptual abilities of the 1-year-old are not at all arbitrary, do not in any way reflect all the possible discriminatory capabilities of the infant (because they no longer discriminate Hindi-only), and are apparently similar to those phonemic categories used by adult native speakers. One possible interpretation of this data is that it may reflect the first stage in the development of a functional phonology. Although more research is needed to determine if this explanation can best account for the reorganization, converging evidence for the emergence of a phonological system around 10 to 12 months has been provided in recent work by Best, McRoberts, and Sithole (1988).

If we accept the definition of a phoneme as a phonetic unit that is used to contrast meaning, then the beginning—just the beginning—of the emergence of a phonemic system around 1 year of age should come as no surprise. Certainly by this time, the infant has begun to construct a receptive vocabulary. Although it is far from clear how fully specified initial oppositions might be, logically the development of a receptive vocabulary and the initial emergence of a system of phonological contrasts should co-occur. We are currently exploring the possible relationship between receptive vocabulary, phonological oppositions, and the reorganization in cross-language speech perception in a series of experiments in our laboratory.

## References

Aslin, R. N., (1987). Visual and auditory development in infancy. In J. D. Osofsky (Ed.), Handbook of infant development (2nd ed., pp. 5–97). New York: Wiley.

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

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

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. Journal of Experimental Psychology: Human Perception and Performance, 14, 345–360.

Bornstein, M. H. (1987). Perceptual categories in vision and audition. In S. Harnad (Ed.), Categorical perception (pp. 287–300.). New York: Cambridge University Press.

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

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

Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. Perception and Psychophysics, 16, 513–521.

Eimas, P. D. (1975). Developmental studies in speech perception. In L. B. Cohen & P. Salapatek (Eds.), Infant perception: From sensation to cognition (pp. 193–231). New York: Academic Press.

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

Flege, J. E. (1984). The effect of experience on Arabs' perception of the English /s/ vs. /z/ contrast. Folia Linquist, 18, 117–138.

Flege, J. E., & Hillenbrand, J. (1986). Differential use of temporal cues to the /s/–/z/ contrast by native and non-native speakers of English. Journal of the Acoustical Society of America, 79, 508–517.

Fodor, J. A. (1983). The modularity of mind. Cambridge, MA: MIT Press.

Hillenbrand, J. (1983). Perceptual organization of speech sounds by infants. Journal of Speech and Hearing Research, 26, 268–282.

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/–/θ/ contrast by francophones. Perception and Psychophysics, 40, 205–215.

Jusczyk, P. W. (1985). On characterizing the development of speech perception. In J. Mehler & R. Fox (Eds.), Neonate cognition: Beyond the blooming, buzzing confusion. Hillsdale, NJ: Erlbaum.

Jusczyk, P. W., Pisoni, D. B., Walley, A., & Murray, J. (1980). Discrimination of relative onset time of two-component tones by infants. Journal of the Acoustical Society of America, 67, 262–270.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. Journal of the Acoustical Society of America, 66, 1669–1679.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. Infant Behavior and Development, 6, 263–285.

Kuhl, P. K. (1985). Categorization of speech by infants. In J. Mehler & R. Fox (Eds.), Neonate cognition: Beyond the blooming, buzzing confusion (pp. 231–263). Hillsdale, NJ: Erlbaum.

Kuhl, P. K. (1987). Perception of speech and sound in early infancy. In P. Salapatek & L. Cohen (Eds.), Handbook of infant perception (Volume 2, pp. 275–382). New York: Academic Press.

Kuhl, P. K., & Miller, J. D. (1975) Speech perception by the chinchilla: Voiced–voiceless distinction in alveolar plosive consonants. Science, 190, 69–72.

Kuhl, P. K., & Padden, D. M. (1983). Enhanced discriminability at the

phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America, 73,* 1003–1010.

Ladefoged, P., & Bhaskararao, P. (1983). Non-quantal aspects of consonant production: A study of reflex consonants. *Journal of Phonetics, 11,* 291–302.

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

Liberman, A. M. (1982). On finding that speech is special. *American Psychologist, 37,* 148–167.

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

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1–36.

Lisker, L. (1985). *Preliminaries to a cross-language study of category naming and discrimination.* Unpublished manuscript.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparitive phonetics. In B. Hala, M. Romportl, & P. Janota (Eds.), *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 563–567). Academia, Prague.

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

MacMillan, N. A., Kaplan, H. L., & Creelman, C. D. (1977). The psychophysics of categorical perception. *Psychological Review, 84,* 452–471.

Mattingly, I. G., & Liberman, A. M. (in press). Specialized perceiving systems for speech and other biologically significant sounds. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Function of the auditory system.* New York: Wiley.

Mehler, J. (1985). Language related dispositions in early infancy. In J. Mehler & R. Fox (Eds.), *Neonate cognition: Beyond the blooming, buzzing confusion.* Hillsdale, NJ: Erlbaum.

Miller, J. L., & Eimas, P. D. (1983). Studies on the categorization of speech by infants. *Cognition, 13,* 135–165.

Pisoni, D. B. (1979). On the perception of speech sounds as biologically significant signals. *Brain, Behavior, and Evolution, 16,* 330–350.

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

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

Stevens, K. N., & Blumstein, S. E. (1975). Quantal aspects of consonant

production and perception: A study of reflex stop consonants. *Journal of Phonetics, 3,* 215–233.

Strange, W. (1987). Speech input and the development of speech perception. In J. F. Kavanagh (Ed.), *Otitis media and child development* (pp. 12–26). Parkton, MD: York Press.

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

Studdert-Kennedy, M. (1986a). Some developments in research on language behavior. In N. J. Smelser & D. R. Gerstein (Eds.), *Behavioral and social science: Fifty years of discovery* (pp. 208–248). Washington, DC: National Academy Press.

Studdert-Kennedy, M. (1986b). Sources of variability in early speech development. In J. S. Perkel & D. H. Klatt (Eds.), *Invariance and variability of speech processes* (pp. 58–76). Hillsdale, NJ: Erlbaum.

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

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

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

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

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

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

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

Williams, L. (1980). Phonetic variation as a function of second-language learning. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology: Vol. 2. Perception* (pp. 185–215). New York: Academic Press.

Younger, B. A., & Cohen, L. B. (1985). How infants form categories. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 19, pp. 211–247). New York: Academic Press.