There are profound changes found across the age range in cross-language speech perception performance, but an adequate understanding of how and why these developmental changes take place is still lacking. In this chapter, I review our research in cross-language speech perception, with emphasis on how it fits into the short history of research in this area. Three overlapping periods of this research will be differentiated.

The first period includes the early seminal work indicating that adults sometimes have difficulty perceiving and producing nonnative phonetic contrasts but that young infants can apparently discriminate nonnative contrasts with ease. The research findings from this period culminated in the mistaken hypothesis that age-related differences in cross-language speech perception result from an absolute loss of perceptual discriminability due to lack of listening experience.

The second period was characterized by a rising skepticism concerning the adequacy of an absolute loss explanation. Research conducted during this period indicated that adults can be trained to discriminate nonnative contrasts, that there are substantial differences in the ease with which nonnative contrasts can be discriminated, even without any training, and that there are significant differences in research findings depending upon the testing procedure employed.

The third period, which represents most of the current research, is characterized by a firm understanding that loss is not an adequate explanation for age and experiential influences on cross-language speech perception and that more complex explanations need to be found. This search for more adequate explanations has led to a rich proliferation of theory-motivated research.
Developmental Changes in Sensitivity to Nonnative Contrasts

When we speak to one another, language is processed with rapidity and ease. However, research in speech processing has shown that this ability is actually quite remarkable. The physical (acoustic) form of any particular utterance, word, or phoneme varies tremendously, depending on the individual speaker, the rate of speaking, and the context in which it is spoken. In order to process speech, the listener must be able to recover the underlying identity over and above this considerable variation. It is of considerable interest to identify the mechanisms that make possible such rapid and efficient processing of speech. Cross-language speech perception research allows a unique perspective on this question by identifying the perceptual abilities of the young infant prior to experience with any specific language and by charting age-related changes in performance as a function of experience with a particular language.

When we began investigating cross-language speech perception, the existing empirical research had indicated that young infants could discriminate both native and nonnative phonetic contrasts (Aslin et al. 1981; Lasky, Syrdal-Lasky, and Klein 1975; Streeter 1976; Trehub 1976), but that adults and children often have difficulty discriminating nonnative distinctions (Goto 1971; Lisker and Abramson 1970; MacKain, Best, and Strange 1981; Miyawaki et al. 1975; Sheldon and Strange 1982; Singh and Black 1966; Snow and Hoefnagel-Hohle 1978; Trehub 1976). On the basis of this fairly consistent pattern of findings, it was hypothesized by several researchers that infants are born with the ability to discriminate the universal set of phonetic distinctions and that this universal ability declines, or is lost, as a function of lack of specific listening experience (Eimas 1975; Strange and Jenkins 1978).

We began our research endeavor in an effort to test the validity of this original hypothesis and, assuming it was true, to establish the point in development at which loss was first apparent. Basic replication research was a necessary first step for two reasons: (1) all of the existing research comparing infants and adults had involved the use of different types of procedures for the two age groups, raising the possibility that the differences in performance between infancy and adulthood stemmed from differences in procedural demands rather than perceptual capabilities, and (2) there were some inconsistencies in the infant data, leading to the possibility that infant phonetic perception might not be as universal as had been claimed.
The most serious concern stemming from the use of different procedures was that the procedures that were used with infants were typically more sensitive than those that had been used with adults. For example, the cross-language infant research involved testing infants in either a high-amplitude sucking (Streeter 1976; Trehub 1976), heart-rate deceleration (Lasky, Syrdal-Losky, and Klein 1975), or conditioned head-turn discrimination task (Aslin et al. 1981; Eilers, Wilson, and Moore 1979). However, many of the adult experiments involved testing subjects in identification tasks (e.g., Lisker and Abramson 1970) or either oddity- or AXB-discrimination tasks (e.g., MacKain, Best, and Strange 1981; Miyawaki et al. 1975).

Identification-, oddity-, and AXB-discrimination tasks all have potentially greater memory demands than the more straightforward discrimination tasks used with infants (for a discussion of the differing demands in adult discrimination tasks, see Carney, Widin, and Viemeister 1977; for a discussion of differential demands in infant tasks, see Eilers and Oller 1988; Jusczyk 1985; Kuhl 1985). It was, therefore, possible that the apparent advantage of young infants was simply an artifact of the more sensitive testing procedures. Thus, the first step was to compare cross-language sensitivity in infants and adults by testing both age groups with the same stimuli and using a similar procedure.

To resolve this problem, we adopted a method of testing which can be implemented in very similar forms with infants (5 1/2 months or older), children, and adults. The procedure used with infants is called the head turn procedure (for a description of this procedure, see Kuhl 1980). Basically, this is a category change task in which the subject has to monitor a continuous background of syllables from one phonetic category (e.g., /ba/) and signal when the stimuli change to a contrasting phonetic category (e.g., /da/). Adults and children signal detection of this change by pressing a button. Correct button presses are reinforced with the presentation of a flashing light for older children and adults or an electronically activated animal for younger children. Incorrect button presses are not reinforced, and misses are not signalled.

The procedure differs only slightly for infants. Infants are conditioned to turn their head toward the sound source when they detect a change in the speech sound. Correct head turns are reinforced with electronically activated animals that become illuminated inside a smoked plexiglass box. As is the case with children and adults, incorrect head turns are not reinforced (for details of this procedure, see Kuhl 1980; for our early implementation, see Werker et al. 1981).
The second potential problem, inconsistencies in the infant data, revolved primarily around contradictory claims with regard to whether young, English-learning infants can discriminate the non-English lead boundary in voice onset time (VOT). In the original infant speech perception experiment conducted by Eimas et al. (1971), it appeared that young infants had the same VOT boundaries as English-speaking adults and that there was no peak at the prevoicing VOT boundary. Similar results were obtained by Butterfield and Cairns (1974), and by Eilers, Wilson, and Moore (1979).

In contrast, Aslin et al. (1981) revealed that English-learning infants do have a sensitivity even to the lead boundary in VOT if tested in a sensitive, staircase procedure. This latter study made it clear that young, English-learning infants can discriminate both native and nonnative VOT contrasts. It also raised the possibility that some contrasts might be perceptually more difficult for young infants than other contrasts, regardless of their status as native or nonnative.

To address the potential interpretative problems arising from variations in ease of discriminability for different contrasts, we decided to measure age-related changes in nonnative speech discrimination using nonnative contrasts that could be expected to vary on their ease of discriminability using linguistic and acoustic criteria. In an early experiment, we compared English-learning infants, aged 6–8 months; English-speaking adults; and Hindi-speaking adults on their ability to discriminate the English- and Hindi-voiced bilabial versus alveolar contrast /ba/-/da/ plus two non-English Hindi contrasts, selected to vary on their potential difficulty (for details see Werker et al. 1981).

The Hindi place-of-articulation distinction between retroflex and dental voiceless stop consonants /Ta/-/ta/ was selected as a potentially difficult non-English contrast as it is rare across the world's languages and has a restricted distribution in those languages in which it does occur (Stevens and Blumstein 1978). The Hindi voicing distinction between breathy voiced and voiceless unaspirated dental stops /d̪a/-/t̪a/ was selected as a potentially easier contrast as it is more common both across and within languages. Also, there was reason to believe that the acoustic cues differentiating the two phones in the retroflex/dental contrast are acoustically less salient than those in the voicing contrast (Ladefoged 1982; Ladefoged and Bhaskararo 1983; Stevens and Blumstein 1978; see Werker et al. 1981 for an analysis of the stimuli).

All stimuli were produced by a native Hindi speaker. Several exemplars of each phoneme were recorded, and eight from each category were se-
lected. Final stimulus selections were based on similarity in intensity, duration, fundamental frequency, and intonation contour.

The results from this early study were as predicted. Virtually all subjects in all groups could discriminate the English /ba/-/da/ contrast. Also, the 6–8 month–old English-learning infants and the Hindi-speaking adults could easily discriminate both Hindi contrasts. However, significantly fewer English-speaking adults could discriminate the Hindi contrasts than either Hindi adults or English-learning infants. English adults had particular trouble with the difficult retroflex/dental place-of-articulation distinction (Werker et al. 1981). Compared to the 100% of Hindi adults who could discriminate both Hindi contrasts, only 40% of the English-speaking adults could discriminate the (potentially easy) Hindi voicing contrast and only 10% could discriminate the retroflex/dental distinction. Following a short training procedure (only twenty-five trials), 70% of the English-speaking adults could discriminate the Hindi voicing contrast, but this training did not improve performance on the retroflex/dental distinction.

The results from this early experiment clarified several points. First, there is an effect of experience on speech perception in that Hindi adult listeners did significantly better than English adult listeners at discriminating both non-English Hindi speech contrasts. The finding that the effect of experience was more pronounced for the retroflex/dental than for the voicing contrast made it clear that some nonnative contrasts are perceptually easier than others. Finally, it was evident that, when using the same procedure, there are significant differences between infants and adults on their ability to discriminate nonnative speech contrasts. Thus these results confirmed and extended the existing data pattern with respect to cross-language speech perception. Infants of 6–8 months of age and Hindi-speaking adults could discriminate both Hindi contrasts with ease, while adult English listeners had difficulty, particularly with the retroflex/dental contrast.

In an attempt to ascertain when in development the change in nonnative sensitivities might first be apparent, we subsequently tested English-speaking children, ages 12, 8, and 4, using the button-press version of the head turn task. Because we viewed the developmental change in cross-language speech perception as mediated by some sort of loss in sensory sensitivity, we expected the decline to be evident around puberty, the age at which Lenneberg (1967) had claimed the critical period closed for acquisition of an accent-free second language (see also Snow and Hoefnagel-Hohle 1977).
To our surprise, the results indicated that children ages 12, 8, and even 4 performed as poorly as English-speaking adults on the Hindi non-English contrasts (Werker and Tees 1983). In fact, the 4-year-old children performed more poorly than the older children and English adults on the easier non-English contrast, the Hindi voicing contrast. This effect was evident even though the 4-year-olds could easily discriminate the English contrast in this procedure and even though Hindi-learning children of this age can also discriminate both of these contrasts when tested in this procedure. These results thus indicated quite clearly that a developmental change is evident before 4 years of age.

It was the poor performance of the 4-year-olds on the Hindi voicing contrast that led us to suspect that the developmental change might be caused by some sort of attentional or perceptual reorganization rather than a sensory loss. This suspicion arose from the fact that 4-year-olds have been shown to be very rigid rule followers in other domains when they have just recently figured out the rules (Carter and Patterson 1982) and may similarly be very rigid rule followers when they first figure out the phonological rules of their native language. Thus, the first seeds of doubt were sown in our minds with respect to the explanation that we and others had given for developmental changes in cross-language speech perception. Nevertheless, before turning to that question, we continued our exploration of when in development there is first evidence of performance decrements in nonnative speech perception tasks.

The next series of studies was begun with two goals in mind. The first goal was to chart developmental changes in cross-language speech perception between infancy and 4 years of age. The second goal was to make sure that the findings with respect to the retroflex/dental distinction would generalize to other nonnative contrasts. In this endeavor, we sought to find another nonnative contrast that would be perceptually difficult for English listeners. It is important to note here that the search was not easy. Many nonnative contrasts are almost immediately discriminable to English adult listeners. However, since we were not interested in pursuing the developmental course of perception of easy non-English contrasts, we continued to search for a more difficult contrast. In this endeavor, we selected a Native American language from the Pacific Northwest because many of these languages have very different phonologies from English. In particular, the phoneme inventory involves an extended series of consonants produced in the back part of the vocal tract, further back than the velar place of articulation used in English.
We selected an Interior Salish language called Thompson, or more accurately, Nthlakampx. This language is spoken by approximately 200 speakers who live around Merritt, British Columbia. Two elders of the community who were known to be excellent speakers and who had some linguistics training (through a program at the University of Victoria designed to help Native Canadian people record and preserve their languages) served as informants. With the help of these speakers, we came up with a list of Nthlakampx words involving minimal pair contrasts. We then recorded the native speakers pronouncing several of these words, followed by a pronunciation of the first consonant and vowel. This method of recording was necessary since there is not an accepted orthography for the language, making it an impossible task to write the CV syllable and making it a nonintuitive task to ask a native speaker to pronounce the CV syllable in isolation.

The next step was to listen carefully to the several recordings. We then selected the syllables that we found to be most difficult to discriminate. The final contrast selected involved a glottalized velar versus a glottalized uvular stop, phonetized by the native informants as /k'į/-/q'į/. As noted in our original paper, the vowels vary somewhat freely in this language, thus several steps were taken to ensure that this consonant contrast could not be detected on the basis of different vowel color (see Werker and Tees 1984a).

English-speaking adults were then compared to both Nthlakampx-speaking adults and English-learning infants on their ability to discriminate the glottalized velar/uvular contrast. As predicted, Nthlakampx-speaking adults and English-learning infants aged 6–8 months could discriminate this contrast, but English-speaking adults showed difficulty (only about 30% could discriminate this contrast).

Following this replication, we began a series of studies attempting to ascertain when in development the decline occurred. Pilot tests involving children between the ages of 8 months and 4 years indicated that changes were taking place around 1 year of life. The pilot work was then followed with a series of cross-sectional and longitudinal studies with infants.

In the first study, English-learning infants aged 6–8, 8–10, and 10–12 months were compared on their ability to discriminate the Hindi retroflex/dental and the Nthlakampx glottalized velar/uvular contrasts (Werker and Tees 1984a). Before being tested on the non-English contrasts, the infants were required to first show they could perform in the Head Turn procedure on the English /ba/-/da/ distinction. All infants were then given twenty-five trials on which to reach discrimination criterion on the non-
English contrast. Before concluding that any infant who failed to reach criterion in that number of trials really could not discriminate the contrast and had not just lost interest in the procedure, all such infants were subsequently retested on the English /ba/-/da/ distinction. The data was only retained as meaningful from infants who subsequently passed the /ba/-/da/ test.

The results indicated that almost all of the infants aged 6–8 months could discriminate both non-English contrasts, but among the infants aged 10–12 months, only 2 out of 10 could discriminate the retroflex/dental contrast and only 1 out of 10, the velar/uvular (Werker and Tees 1984a). The infants aged 8–10 months of age showed an intermediate pattern of performance. This pattern was replicated in a longitudinal study in which a small group of six infants were tested at two-month intervals (Werker and Tees 1984a). Using the same procedure, a few infants who were learning Hindi and/or Nthlakampx as a native language were shown to be able to discriminate the contrast from their language-learning environment when they reached 11 months of age (Werker and Tees 1984a).

More recently, the results with respect to the English-learning infants were replicated in the same procedure using synthetically produced voiced retroflex/dental stimuli (Werker and Lalonde 1988). English-learning infants aged 6–8 months were shown to be able to discriminate multiple, synthetically produced tokens according to the adult Hindi retroflex/dental boundary but not according to an arbitrary boundary location that does not correspond to any adult phonetic category. This result with the younger infants showed that infant sensitivity to nonnative phonetic categories is categorical-like and is related to the phonetic relevance of the contrast in question. When given the same number of testing trials, English-learning infants aged 11–13 months were not able to discriminate the stimuli according to either the arbitrary boundary location or the Hindi retroflex/dental boundary but were, of course, able to discriminate them according to the English ba/da boundary.

Also, Best and McRoberts (1989; Best, this volume) have replicated the developmental change between 6 and 12 months of age for English-learning infants tested on the Nthlakampx glottalized velar/uvular contrast, /k\̣ːi/-/q\̣ːi/, using a habituation/dishabituation testing procedure rather than the head turn procedure used in our previous work. Taken together, these replications with new contrasts and different testing procedures provide strong confirmation that the developmental change in nonnative speech perception evident within the first year of life is related to listening experience.
Sensitivity to Nonnative Contrasts

During the second period in cross-language speech-perception work, researchers became increasingly skeptical about the appropriateness of the strong claim that experiential effects are equally apparent for all nonnative contrasts and that experiential effects are permanent. Research made it clear that the effect of experience does not equally affect all nonnative contrasts as adults can discriminate many nonnative contrasts with little difficulty (see, for example, Eilers, Wilson, and Moore 1979). Similarly, our research showed that the Hindi voicing contrast is easier than the place-of-articulation contrast for English listeners (Werker et al. 1981). The finding that minimal training can improve performance on at least some nonnative contrasts strengthened the possibility that experiential influences might not be permanent. Questions still remained as to whether it is possible to train adults to discriminate all nonnative distinctions, whether some training methods are more effective than others, and whether some nonnative contrasts are untrainable. A careful consideration of the historical context in which the work was done helps trace how the understanding of these issues has changed over the years.

One of the most influential early training studies was conducted by Winnifred Strange (1972). In this study, discrimination training was shown to have limited effectiveness on English adults’ ability to discriminate non-English VOT contrasts. Because this finding was consistent with the expectations generated by the seminal work of Lisker and Abramson (1967; Abramson and Lisker 1970), it was largely unchallenged by the research community for a number of years. Subsequently, however, a number of investigators have reported findings indicating that English speakers can be fairly easily trained to discriminate the non-English lead boundary in VOT. For example, Carney, Widin, and Viemeister (1977) were able to train English adults to discriminate synthetic stimuli differing in VOT at a number of arbitrary points along a continuum using a same/different (AX) procedure with a short ISI and feedback. Moreover, Pisoni et al. (1982) reported that English adults’ ability to discriminate the lead boundary in VOT could be improved by simple labeling of prototypical stimuli from three VOT categories. Such minimal training even generalized to VOT distinctions at a novel place of articulation (McClaskey, Pisoni, and Carrell 1983).

This kind of research allowed the field to move away from the question of whether or not training works to the more complicated question of why some training procedures are more effective than others. The training
procedure employed by Strange in the original 1972 study involved primarily discrimination training, whereas that employed by Pisoni and his colleagues (1982) required subjects to label the stimuli.

Perhaps, as suggested by both Jamieson and Morosan (1989) and by Pisoni, Lively, and Logan (this volume), training procedures involving labeling are more effective at facilitating linguistically relevant perception. Consistent with this notion, Strange and Dittman (1984) reported little success at training Japanese adults to discriminate /r/ from /l/, and MacKain, Best, and Strange (1981) found that only extensive, naturalistic second-language learning was effective at improving Japanese adults' ability to discriminate the English /ra/-/la/ distinction. However, Logan, Lively, and Pisoni (1991) report that, in some training procedures, Japanese adults can learn to discriminate the English /r/-/l/ distinction. In this study, Logan and colleagues trained Japanese listeners on the English /r/-/l/ distinction over a period of three weeks, using a two-choice labeling procedure. They ensured that the subjects were trained in a variety of contexts (different position in syllable) and that they were exposed to multiple speakers. Thus, training conditions were set up to facilitate generalization.

It is useful to note that, even when training is quite successful, adults may fail to achieve nativelike levels of performance (see Polka 1991). For example, in their recent study, Logan, Lively, and Pisoni (1991) note that, although the overall change in performance between the pre and the posttest sessions was highly significant, the overall amount of improvement was less than 10%, even though training continued for three weeks. Also, the relative success of training varied according to position in the syllable (it was not at all successful for clusters in initial positions), and training generalized more to similar than to dissimilar testing contexts. Finally, even after this amount of training, it is questionable as to whether Japanese adults were performing as well as their English-speaking counterparts.

There is some recent research showing that nonnative listeners show more difficulty perceiving even relatively easy phones than do native listeners under certain testing conditions. In recent work, Takata and Nabelek (1990) compared native English speakers to native Japanese speakers who are fluent in English on their performance in the modified Rhyme Test. Results indicated that, although the two groups performed similarly under quiet testing conditions, the native Japanese speakers performed significantly more poorly than the native English speakers in conditions of noise and/or reverberation. Not surprisingly, one of the more common errors for native Japanese listeners was an r/l confusion.
Thus, there is no doubt that linguistic experience has a profound effect on speech perception, but there is also no doubt that the effects of experience can be ameliorated in certain training and testing situations. Perhaps the most useful way to present the results of cross-language training studies is to report not only the amount of improvement between pre and posttest training but also to report the results of analyses comparing post-training performance levels to those of native speakers (see Polka 1991).

It is also useful to assess the long-term effectiveness of training. For example, in their work, MacKain, Best, and Strange (1981) evaluated the effectiveness of second-language training on the long-term consequences of discrimination. In related work, our early research showed that training immediately facilitated performance on the Hindi voicing contrast /dʰa/-/tʰa/ (Werker et al. 1981) and was not successful at facilitating performance on the retroflex/dental contrast, /ta/-/Ta/. In a subsequent study, we found that more extensive discrimination training (500 trials) did significantly facilitate performance of at least some English speakers on the retroflex/dental contrast. However, the effect of training had disappeared when subjects returned to the lab a few weeks later (Tees and Werker 1984). In this experiment, training was clearly not as sophisticated as that used in other work (Logan, Liverly, and Pisoni 1988; Strange and Dittman 1984). The training trials simply involved feedback at regular intervals during testing in the button-press category change procedure. It is quite possible that an alternative training procedure would have been more effective at improving long-term performance. ²

The focus on loss as an explanation for the attenuation of discrimination performance resulted in the ignoring an important fact, namely, poor discrimination performance was rarely all-or-none even without any training. Some residual auditory discrimination skill remains. Thus, it seemed possible that listeners have multiple means to the same ends.

In previous work, it had been shown that adults may use both auditory and phonemic processing in their attempts to discriminate sounds (see also Repp 1984). To examine whether adults would show a similar sensitivity to nonnative perception even without training, we tested English adults on the Hindi retroflex/dental and the Nthlakampx-glottalized velar/uvarular contrasts in a more sensitive procedure. Because we were interested in assessing sensitivity, we tested adult English speakers in a same/different (AX) discrimination task (Carney, Widin, and Viemeister 1977). Using the AX procedure, we found that adult subjects can discriminate both contrasts at a 500 but not a 1500 msec ISI (Werker and Tees 1984b).
In a subsequent study using just the Hindi retroflex/dental stimuli, we tested subjects for five blocks of trials in one of three ISI conditions, 1500, 500, and 250 msec (Werker and Logan 1985). Again, the results revealed sensitivity to the nonnative phonetic contrasts in the shorter ISI conditions. In fact, there was even evidence that subjects can discriminate nonphonetic acoustic cues within either the retroflex or dental category at the 500 msec ISI (Werker and Logan 1985).

In an attempt to make sense out of this pattern of findings, we proposed that subjects can use one of three different processing strategies—phonemic, phonetic, and acoustic—depending on the interstimulus interval. When tested with an ISI over 500 msec, subjects appeared to use a phonemic processing strategy and were unable to discriminate the nonnative contrast. Thus, when the ISI is long, subjects seem unable to discriminate two stimuli unless they can assign them distinct linguistic labels. At shorter ISIs, subjects showed evidence of using both a phonetic and an acoustic strategy. Evidence for a phonetic strategy was provided by subjects who could discriminate retroflex from dental exemplars but could not discriminate among the several exemplars within either phonetic category. Evidence of acoustic processing was provided by subjects who could discriminate between the several retroflex or the several dental exemplars. These findings indicate that adult listeners can discriminate between tokens on the basis of phonetic and acoustic information if the task requires it but that the most readily available strategy is to perceive speech stimuli in terms of native-language phonemic categories.

This pattern of results has been replicated using a new contextual manipulation and using synthetic rather than naturally produced retroflex and dental tokens (Morosan and Werker, in preparation). The synthetic tokens were constructed by varying the starting frequency of the second and third formats in equal steps, thus ensuring that the acoustic variability within categories was equivalent to that between categories (for stimulus descriptions, see Werker and Lalonde 1988). The contextual manipulation in this study involved varying the kinds of pairings used in the stimulus set rather than manipulating ISI or the number of trials as we had done in our previous work.

There were three contextual conditions in this study, with ten subjects in each condition. In the phonemic contextual condition, there were four kinds of pairings: phonemically different (bilabial/dental), phonetically different (retroflex/dental), acoustically different (two different bilabial, dental, or retroflex stimuli), and physically identical (the same token paired with itself). There were equal numbers of one- and two-step pair-
ings among the phonemically different, phonetically different, and acoustically different trials. In the phonetic contextual condition (phonemic in other languages but not in the listener's language), the pairings that are phonemic to English listeners were eliminated. In the acoustic contextual condition, only the within-phonetic category (acoustically different) and physically identical pairings were presented.

All subjects received two blocks of ninety-six trials each. Responses to phonemically, phonetically, and acoustically different pairing types were converted to A' scores using performance on the physically identical pairings to estimate false alarm rate. As expected, subjects performed nearly perfectly on the phonemically different pairings in the first contextual condition. Of more theoretical interest is the relative performance on phonetically different and acoustically different pairings across the contextual manipulations.

If the contextual manipulation of pairing type affects speech perception, it could have an effect in at least three different ways. One possibility would be that subjects attend primarily to the largest acoustic difference present within a contrast type and ignore other smaller acoustic differences: subjects would attend to the two-step pairings in each condition and ignore the one-step and physically identical pairings. Thus, according to this prediction, the proportion of different responses to phonetically different and acoustically different pairings would not change across the three testing conditions, it would simply always reflect a bias in favor of larger acoustic differences irrespective of phonetic status. There was no consistent support for this prediction.

A second possibility is that subjects attend to the most easily accessible linguistic difference. According to this prediction, subjects would show above-chance discrimination of only phonemically different pairings in the first contextual condition, above-chance discrimination of only phonetically different pairings in the second contextual condition, and would only show above-chance performance on acoustically different pairings in the third contextual condition when neither phonemically or phonetically different pairings were present. The data pattern did not fit this prediction.

The third possibility is that the presence of phonemically different pairings would enlist the linguistic mode. In this case, the perception of phonetically different pairings would be facilitated in the first contextual condition relative to the other two conditions due to the presence of pairings that have functional phonological status in English in the first contextual condition. In the subsequent two contextual conditions, with no pairings present that have functional phonological status in English,
the English listeners would process both the phonetically different and acoustically different pairings according to a nonlinguistic, acoustic strategy. As shown in figure 4.1, the results were consistent with this prediction.

These results support the hypothesis that the presence of phonemically different pairings engages a linguistic mode of processing, thus facilitating sensitivity to pairs of stimuli that linguistically straddle phonetically relevant boundaries. These results replicate the previous finding that indicates that adults can use at least three different processing strategies—phonemic, phonetic, and acoustic—depending upon testing conditions (Werker and Logan 1985; Mann 1986). It also clarifies, however, that the phonemic mode has the most privileged status.

In summary, I have characterized the second period of research in cross-language speech perception as reflecting a growing realization that developmental changes do not result in permanent loss and that careful attention to the testing situation is necessary to understand disparate results.

**Theory-Guided Research in Cross-Language Speech Perception**

More recently, much of the research in cross-language speech perception has moved beyond simple demonstrations of whether subjects can or cannot discriminate particular nonnative contrasts to an attempt to test,
within a theoretical framework, specific predictions, such as ease of discriminability and ease of training. For example, several researchers have attempted to make theoretically based predictions specifying which nonnative contrasts will be easy or difficult to discriminate (Best, McRoberts, and Sithole 1988; Burnham 1986; Strange 1986).

Burnham (1986) suggested that there might be both fragile and robust nonnative distinctions. Fragile refers to phonetic contrasts that are both rare across the worlds' languages and, of particular importance, are acoustically quite similar. Burnham hypothesized these types of contrasts to be most vulnerable to early loss—around 10–12 months of age—in nonnative listeners. Robust refers to contrasts that are widely distributed across the world's languages and are acoustically less similar. Burnham hypothesized that nonnative listeners would not show a measurable decline in their performance on these contrasts until around 4–5 years of age.

Because Burnham viewed acoustic salience as the most important dimension of fragile contrasts, he reasoned that early loss of fragile contrasts would result from sensory tuning (e.g., Aslin and Pisoni 1980; Gottlieb 1981). He posited a different mechanism for later loss of robust contrasts. Specifically, he argued that, with the advent of metaphonological abilities, sensitivity to even robust contrasts would decline because the distinctions have no functional value in the individual's working phonological system.

Although the experimental data that Burnham presented in support of late loss for robust contrasts was less than convincing, the idea that some types of contrasts may simply be less vulnerable to loss at any age than others was immediately seized upon by the research community as an important idea. Also, his suggestion that more than one explanation for developmental changes in cross-language speech perception might be required is of interest.

Catherine Best and her colleagues have taken the strong stand that phonological status alone should predict whether a contrast is discriminable or not to a nonnative listener. Because this research is covered in depth in the chapter by Best (this volume), it will only be touched on here. Basically, Best, McRoberts, and Sithole have proposed that there are at least four kinds of nonnative contrasts in terms of phonological status: (1) assimilable, (2) nonassimilable, (3) category goodness, and (4) two category. Assimilable nonnative contrasts are those in which each member of the contrast can be assimilated to an intermediate phone in the native language. These kinds of nonnative contrasts should be the most difficult
to discriminate. Nonassimilable contrasts include phones that do not even sound at all like any possible phone from the native language. Because these phones do not invoke phonological processing, they should continue to be discriminable by basic auditory processes throughout the life, and as such, these nonnative contrasts are predicted to be the most easily discriminable. Category goodness refers to a nonnative contrast whose members can each be assimilated to an intermediate phoneme in the native language, as in assimilable, but one which will stand out as clearly a better instance of that category than the other. Two category refers to a nonnative contrast that consists of two nonnative phones, each of which is assimilable to a contrasting phonemic category in the native language. The third type is predicted to be intermediate in difficulty to purely assimilable and two-category nonnative phones; the last type is predicted to be easiest.

Research to date is consistent with these predictions. In a series of studies, Best and colleagues have shown that Zulu click contrasts that are not at all assimilable to English are easily discriminated by English-speaking subjects of all ages, including 12- to 14-month-old English-learning infants (Best, McRoberts, and Sithole 1988). In subsequent work, Best and colleagues have tested both adults and infants of different ages as to their ability to discriminate other non-English contrasts. In the case of adults, the relative difficulty of discrimination can be fairly well predicted by the phonological status of the contrasts (Best 1989). However, the results with respect to infants are a bit more confusing (Best, this volume; Best and McRoberts 1989; Best et al. 1990). Infants of 6–8 months of age clearly perform better than infants of 10–12 months of age on all but the nonassimilable contrasts, but beyond that, the pattern of performance does not follow that predicted by Best's phonological assimilation model, suggesting that, in the older infants, perception is not yet organized by the same phonological constraints as it is in adults.

In recent work, Polka (1991, 1992) has highlighted at least three independent factors that need to be considered when making predictions concerning the discriminability of nonnative contrasts among adults. These are functional phonetic status (phonemic contrast), substantive phonetic status (phonetic variation), and acoustic differences (the absolute amount of measurable acoustic differences between members of a nonnative contrast irrespective of phonetic status) (see also Best, this volume). She has argued that all three of these factors need to be considered in assessing the discriminability of a nonnative contrast for subjects of any age.
In one study, English- and Farsi-speaking adults were compared as to their ability to discriminate the Nthlakampx glottalized velar/uvular contrast used in our previous work (Polka 1992). Glottalized stops do not have substantive phonetic status (they do not occur) in either English or Farsi. However, the Farsi language does include a velar/uvular functional phonetic (phonemic) contrast between velar and uvular for nonglottalized stops. If perception of nonnative contrasts is predicted by their match to the system of phonemic possibilities in the language irrespective of their phonetic substantiation, then Farsi speakers should find this contrast easier than English speakers. On the other hand, if the phonemic contrast has to be supported in the identical phonetic environment in a language, then Farsi speakers will be no better than English speakers.

Subjects were tested in an AX procedure with a long (1500 msec) ISI. Although there were no significant differences between the English and Farsi speakers in overall performance, there were substantial individual differences. The Farsi speakers who could hear the Nthlakampx stimuli as peculiar Farsi sounds did better on the glottalized velar/uvular contrast than the Farsi speakers who did not recognize the stimuli as similar to Farsi.

Polka and colleagues have extended this line of investigation to testing English adults on their ability to discriminate the Hindi retroflex/dental contrast as instantiated in four categories of voicing. This contrast does not have functional phonetic status in English. In this particular case, both substantive phonetic characteristics and a metric of acoustic discriminability were found to be important, providing at least some support for the notion that each contributes to discriminability among adults (Polka 1991, 1992).

Flege's work on speech perception and production among second-language learners constitutes another example of theoretically motivated research in this area. His goal has been to identify the extent to which nonnative contrasts share acoustic and/or articulatory cues with native distinctions and whether such overlap predicts ease of acquisition of productive and perceptual abilities (for a review, see Flege 1992).

In this endeavor, Flege tests specific theoretical predictions about the kinds of errors second-language learners will make in producing and perceiving L2 (second language) phones. He has shown that, when first acquiring a new language, subjects more rapidly attempt to pronounce phones that are similar to those used in their native language, although they make mistakes and pronounce the new phone as if it were identical to the similar phone in the native language. Adult second-language learn-
ers will avoid even attempting unfamiliar sounding phones in the early stages of second-language acquisition (Flege 1987; see also Wode 1977, 1992). However, at a subsequent point in the acquisition process, they continue to mispronounce the similar phones, presumably because they continue to assimilate them to native categories, but they become better at correctly pronouncing the dissimilar phones.

The data are still somewhat ambiguous as to whether the ultimate ability to pronounce accurately dissimilar phones stems from the establishment of a new underlying phonetic category or whether it stems from the use of the same underlying phonetic representation with the application of different language-specific realization rules (Flege 1992). Of interest, Flege has clear data to suggest that young L2 learners can set up new underlying representations for either similar or dissimilar L2 phones. His research is now directed at resolving the ambiguity of the explanation with respect to adults and to extend systematically the model to perception studies. His work is characterized by a systematicity that should ultimately allow us to resolve the ambiguity of the extent to which notions of similarity and assimilability can be most usefully understood with reference to acoustic, articulatory, or phonological properties.

We recently outlined several different approaches that can be used to explain experiential influences on cross-language speech perception (Werker 1991; Werker and Pegg 1992). These included perceptual tuning, cognitive mediation, phonological processing, modular recalibration, self-organizing systems, and articulatory mediation. Note that most of these postulated mechanisms involve processes other than sensory loss. Our current research is designed to attempt to evaluate the relative utility of these various kinds of explanations and ultimately to determine if a single process or combination of several different processes is required to understand age- and experience-related influences on cross-language speech perception (Lalonde and Werker 1990, under review; Werker and Pegg 1992).

To date, we have been concentrating our efforts on testing the viability of the cognitive mediation and phonological processing alternatives. We have recent data to suggest that the ability to restructure perceptual categories for visual information on the basis of correlated attributes (Younger and Cohen 1983) emerges in tandem with the developmental reorganization in cross-language speech perception (Lalonde and Werker 1990, under review), raising the possibility that both rely on the same underlying cognitive prerequisites. This kind of data may be consistent with the attentional explanations for developmental changes in speech.
Cross-Language Speech Perception

perception postulated by Jusczyk (1992, this volume). However, we are not entirely convinced that cognitive mediation necessarily implies that speech perception is accomplished by general purpose cognitive machinery. Thus, we have also continued our investigation of the relationship between developmental changes in cross-language speech perception and the development of other specific linguistic abilities.

Our research testing the viability of the phonological processing alternative was originally motivated by the three-factor model of adult speech perception we proposed (Werker and Logan 1985). We reasoned that if adults can use phonemic, phonetic, or acoustic processes in perceiving speech, then it should be possible to trace to developmental emergence of these three factors and that the emergence of the phonemic factor should signal the onset of phonological processing. Toward this end, we conducted two sets of experiments which will be reviewed below.

In the first set of experiments, Lalonde and I replicated and extended the Werker and Tees (1984a) finding of a developmental reorganization between 6 and 12 months of age, but we substituted a synthetically produced /ba/-/da/-/Da/ continuum for the natural stimuli used in our previous work. The synthetic stimuli were necessary to control for the confound between physical similarity and phonetic status inherent in natural stimuli. English-learning infants aged 6–8 months were compared to English-learning infants aged 10–12 months as to their ability to discriminate six stimuli from either side of three locations along this continuum. The first location was labeled Common because it contrasted three bilabial /ba/ with three dental /da/ stimuli—a phonetic contrast common to both English and Hindi adults. The second location was labeled Hindi-only because it contrasted three dental /da/ with three retroflex /Da/ stimuli as judged by Hindi-speaking adults (see experiment 1, Werker and Lalonde 1988). The third location was labeled Neither as it required subjects to treat a dental and two retroflex stimuli as equivalent to one another and as different from three additional retroflex stimuli. Thus, infants were asked to discriminate the stimuli according to a location that has no phonetic relevance.

The results are shown in figure 4.2. As can be seen, both groups of infants were able to discriminate the common contrast, only the 6- to 8-month-old infants could discriminate the Hindi-only contrast, and infants in both age groups were unable to discriminate the Neither contrast. In terms of the three factors (phonemic, phonetic, and acoustic) proposed in our previous work, we interpreted these results as revealing a phonetic, or at least, phonetically relevant, factor in the performance of the infants
aged 6–8 months since they were able to discriminate the non-English Hindi-only contrast, and a phonemic factor as evident in the sensitivity to only the common contrast among the infants aged 11–13 months of age.

There was no evidence for an acoustic factor in this study. On the basis of previous work by Aslin and colleagues (Aslin et al. 1981), it is suspected such evidence would be apparent if infants were tested in a more sensitive, staircase procedure. In other words, it seems logical to expect that infants will also have general acoustic-processing sensitivities that might be evident under certain experimental testing conditions. In fact, it is possible that infants aged 11–13 months might even show continued sensitivity to the non-English Hindi-only contrast if tested under adequately sensitive conditions.

Further research is needed to test both of the above predictions. Nevertheless, on the basis of the results from the Werker and Lalonde (1988) study, there is solid evidence for performance in accord with both pho-

Figure 4.2
Percentage of correct responses by group for each kind of pairing.
nnetic and phonemic factors at different points in infancy. The phonetic factor is evident in the performance of the infants aged 6–8 months, and the phonemic factor, in the performance of the infants aged 10–12 months.

Because we were attempting to understand the infant data in terms of the three-factor model outlined with adults and had consequently labeled the performance of 10- to 12-month-old infants as phonemic, we postulated that infants of 10–12 months of age should have the beginnings of phonemic oppositions in the representation of lexical items in their receptive vocabularies. We thus began a series of experiments designed to test this possibility (for details, see Werker and Baldwin 1991; Werker and Pegg 1992).

However, two years of experimentation failed to reveal any convincing evidence that infants represent words in such fine detail by this young age. In fact, the first age at which we have replicable evidence of sufficient phonetic detail to enable minimal pair contrasts in the receptive lexicon is 19 months (Werker and Baldwin 1991). Thus, it is probably inaccurate to refer to the reorganization at 10–12 months of age as involving phonemic processing since “phonemic” implies the ability to use phonetic detail to contrast meaning. Therefore, we now feel it is probably more accurate to characterize the performance of the infants aged 10–12 months as involving language-specific phonetic sensitivities.

On the basis of this theory-motivated research, we have now modified the previously proposed three-factor model of speech perception (Werker and Logan 1985) and replaced it with a four-factor model, acoustic, broad-based phonetic, language-specific phonetic, and phonemic factors (for a more complete description, see Werker and Pegg 1992). As elaborated in Werker and Pegg, we are convinced that there is evidence of at least two factors in infancy. One is the language-specific phonetic perception seen by 10–12 months of age, and the other is the broad-based phonetic sensitivity of the younger infant. To date, there is no clear evidence showing if there is an independent acoustic factor in infant perception, but as mentioned above, there is reason to believe such evidence might be found. We feel the previous work by Werker and Logan (1985) and by Morosan and Werker (in preparation) provides clear evidence for at least three factors in adult speech perception: acoustic, broad-based phonetic, and phonemic. To date, however, there is no data that might differentiate language-specific phonetic processing from phonemic processing among adults.
Finally, it is interesting to note that the study of cross-language speech perception has now extended far beyond the study of phonetic and phonological perception, to investigations of the perception of global prosody (Mehler et al. 1988), phrasal and clausal structure (Jusczyk 1989), and phonotactic rules (Jusczyk, Friederici, and Wessels, in press). Much of this work is motivated by a realization that the understanding of phonetic perception is intimately tied up with an understanding of how words and segments are extracted from ongoing speech (e.g., Jusczyk, this volume; Mehler, Dupous, and Segui 1990; Pisoni and Luce 1987).

Conclusion

In this chapter, I have attempted to review our work in cross-language speech perception within a historical perspective. I hope it is apparent from this selective review that cross-language speech perception represents a dynamic research area in which the ideas have grown and changed over the years. In an attempt to illustrate just how much our thinking has progressed in this research area, three periods in cross-language speech perception research have been identified. The first period was characterized by the view that language experience is necessary to maintain the ability to discriminate nonnative contrasts and that, without such experience, the ability to discriminate phonetic contrasts will be lost. This was, of course, an overly simplistic view: even in his original statements, Gottlieb (1976) pointed out that maintenance at the behavioral level may not be the same thing as maintenance at the neuronal level (see Walley, Pisoni, and Aslin 1981; or see Aslin, Pisoni, and Jusczyk 1983 for an elaboration of how attunement theory can allow for either sensory or attentional mechanisms).

In the second period, it became apparent that there are substantial differences between nonnative contrasts in their discriminability and that subjects can discriminate even difficult contrasts if given adequate training or if tested in a sensitive enough procedure. This led to a realization that sensory loss was not an adequate explanation and that, at the very least, developmental changes in cross-language speech perception should be characterized as involving a reorganization in perceptual biases rather than a loss in absolute discriminatory abilities (Werker and Tees 1984a).

Current work in cross-language speech perception is increasingly theory motivated (see MacKain 1988). As noted above, researchers are now designing experiments to test specific predictions generated from different
theoretical perspectives. This development reflects an increasing sophistication in the field, and will, we can hope, lead to a more adequate understanding of developmental changes in cross-language speech perception.

Notes

1. Judgments of difficulty were arrived at in several steps. First, I listened to the syllables and selected several that I found difficult to discriminate. I then played those to a group of faculty members in linguistics at UBC and selected the contrasts that everyone but the trained phoneticians was unable to discriminate. I then digitized and analyzed those several syllables at Haskins Laboratoris and played them to colleagues and students at Haskins. The final set of syllables selected was judged to be the most difficult by those listeners.

2. It is of interest to note that English-speaking adults who had early exposure to Hindi during the first couple of years of their life but no subsequent systematic exposure did significantly better than the native English speakers on this contrast. Also, all effects of training were permanent with this group.

3. Notice that this whole area of research moves beyond that investigated in our early work. We did not attempt to identify what kinds of contrasts might be difficult or easy; we simply tried to find difficult nonnative contrasts in order to test our hypotheses.

4. For recent research consistent with an articulatory mediation hypothesis, see de Boysson-Bardies et al. 1992.

References


Cross-Language Speech Perception


