Chapter • 1

Cross-Language Studies of Speech Perception *A Historical Review*

Winifred Strange

The research programs reported by the contributors to this volume address questions about the role that linguistic experience plays in shaping the way speech is perceived by infants, children, and adults. In other words, the main question of interest is how the perception of spoken language is influenced by language learners' history of interactions with their linguistic environment (i.e., the community of people who know and use a particular language).

Human beings are biologically endowed with remarkable sensory, motor, and cognitive capacities that enable them to learn to communicate by linguistic means. The primary mode through which speakers and listeners communicate linguistic messages is via the "speech chain" (oral/aural system). Speakers render their linguistic intentions into sequences of speech movements (articulatory gestures). These, in turn, generate complex acoustic signals that are picked up via the auditory system of the listener and "interpreted" so that the linguistic intention is recovered. Over the course of the first 5 to 8 years of life, almost all children with normal cognitive and sensory functions learn the basic structure of their native language (i.e., the language of their caregivers). Furthermore, human beings are capable of learning additional languages during their lifetime, given sufficient experience with, or formal instruction in, the non-native language(s). Thus, the human capacity to learn language via experience with spoken language input is maintained throughout the life span.

In this book, we are concerned with only one aspect of spoken language learning; that is, learning to perceive and produce the phonological structures of the language. More specifically, the primary focus is on how listeners perceive the sequences of "speech sounds" (consonants and vowels) that make up the syllables and words of a language. For purposes of discussion of the many research programs

that explore experiential effects on speech perception, the book is divided into three sections.

1. investigation of how speech perception develops in the course of learning our first (native) language
2. assessment of how patterns of speech perception may change when we learn a subsequent (foreign) language
3. exploration of how speech perceptual patterns may be modified in the laboratory or clinic by manipulating the listening experiences of subjects or clients

The goal of this chapter is to provide a conceptual framework within which the ideas and findings presented in subsequent chapters can be considered. To accomplish this, a selective history of research in cross-language speech perception is provided. In this summary, theoretical themes that have motivated research in speech perception are presented, methodological paradigms that have dominated the field are described, and some of the most important empirical findings of the last 15 years are discussed. However, before this review, the next section sets the stage by presenting a brief description of the basic phenomena that serve as starting points for the investigation.

PRELIMINARIES

The Constancy Problem in Speech Perception

Although research on human perception is as old as psychology itself, empirical study of the perception of speech is of relatively recent origin. Using an analysis-by-synthesis strategy made possible by the invention of the sound spectrograph (Potter, Kopp, and Green 1947) and the acoustic speech synthesizer (c.f. Cooper 1950), researchers explored the "proximal stimulus" for speech perception, that is, the complex acoustic patterns from which perceivers recover the linguistic message communicated through the talker's gestures. Early researchers soon discovered that there was no simple correspondence between segments of the acoustic signal on the one hand and perceived units as they were characterized by phoneticians/phonologists on the other. The physically continuous acoustic signal could not be segmented temporally into distinctive entities that corresponded uniquely to the units of a phonetic transcription. Thus, early theorists described the acoustic patterns associated with spoken utterances as a phonetic "code" rather than a "sound alphabet" (Liberman et al. 1967).

Speech perception, then, provides an example of one of the ubiquitous problems in perception, that of perceptual constancy. Humans perceive objects and events in the environment as belonging to categories. In the case of speech, one set of categories of interest are phonetic categories, the smallest segments of spoken language that combine and contrast to make up the words of the lexicon. In linguistic analysis, these distinctive phonetic categories are called phonemes. The problem of perceptual constancy arises because there is no one-to-one correspondence between phonemes as perceived and the acoustic patterns generated by speech gestures that constitute the stimuli for speech perception. Thus, many physically different acoustic patterns may be categorized as the same phoneme (many-to-one correspondence), and even more inexplicably, the identical acoustic signals are often categorized as different phonemes (one-to-many correspondence) when they occur in different contexts, or in utterances produced at different speaking rates or spoken by different talkers.

In the face of this lack of invariance in the speech stimulus, a basic question to be addressed is how humans come to be able to categorize acoustic patterns correctly. That is, how do listeners recover the phonetic segments intended by the talkers, thereby achieving perceptual constancy?

Classical answers to the perceptual constancy problem (postulated primarily to explain such visual phenomena as constancy of the size and shape of three-dimensional objects) included two basic theoretical positions: 1) perceptual categories are learned through association of inherently ambiguous proximal stimuli (e.g., the retinal image) with other experiences with the distal objects and events (associative learning position); and 2) perceptual constancy is a function of innately given mental categories (nativist position). A third alternative offered by James J. and Eleanor J. Gibson rejects the basic premise of the first two positions, that the stimulus is inherently ambiguous. According to their direct realist position, there is sufficient information in stimulation (the patterning of energy over time) to specify the perceived objects and events. According to this theory, the apparent ambiguity of the stimulus comes from an inappropriate level of analysis of the physical environment. Veridical perception of real objects and events is achieved because perceivers (learn to) detect the invariant patterns in the energy array that uniquely specify those objects and events (c.f. Gibson 1979; Gibson 1992).

In both early and current theories of speech perception, we can see the influence of each of these theoretical positions. Early versions of the motor theory (Liberman et al. 1967) clearly fell within the tradition of associative models. In contrast, feature detector models (Eimas and Corbit 1973) posited biologically determined (innate) mechanisms by which phonetic categories were differentiated. More recently, revised motor theory (Liberman and Mattingly 1985) combines aspects of both associative and nativist positions. Biologically given

"modules" (cognitive/neural mechanisms) are specialized to process speech stimuli; invariant phonetic segments are recovered from the variable acoustic structures by reference to underlying gestural specifications. Finally, direct realist approaches to the problem of understanding speech perception have been formulated (Fowler 1986). (See Best, this volume, for further discussion and comparisons of current theories of speech perception.)

Units of Analysis

In the description of the constancy problem just presented, the "objects" of speech perception (i.e., perceptual categories) were characterized in terms of linguistic units of analysis, the most basic of which is the phoneme. In exploring questions about the perception of speech in general and the role of experience in speech perception in particular, it is necessary to question that characterization. We must ask anew how the objects of perception might best be characterized and then ask how those characterizations relate to descriptions of acoustic signals as stimuli and to the hypothetical mechanisms by which the objects of perception are apprehended from the acoustic input. Specifically, it is important to consider the traditional distinction between phonetic and phonological (or phonemic) levels of analysis of spoken language and how each relates to speech gestures as spoken and acoustic signals as heard.

In *phonetic* analysis, submorphemic segments (phones) are characterized primarily in terms of their substantive articulatory properties. The Universal Phonetic Inventory (captured by the International Phonetic Alphabet) describes a structured set of segments that occur in the languages of the world. Relationships among segments of this inventory are defined by (abstract) articulatory features that capture their similarities and differences. Consonants (which involve one or more constrictions of the vocal tract) are classified with respect to their manner of articulation (degree of constriction), their place of articulation (location in the vocal tract of the constriction), and sometimes the shape of the articulators (primarily the tongue). Additional articulatory features include voicing and aspiration (degree of laryngeal restriction), nasality (position of the velum), and features describing airstream dynamics (ingressive vs. egressive, ejective vs. implosive). Vowels (which are produced with a relatively open vocal tract) are described in terms of tongue height and backness (position of the tongue body in the oral cavity), position of the jaw, posture of the lips, and length of the gesture.

In a *phonological* analysis, the primary aim is to establish how classes of phones combine and contrast to form the distinctive words of a particular language. That is, phonetic segments are characterized

in terms of their linguistic function. The phoneme inventory of a particular language is the set of abstract phonetic categories that are perceived to be different by native speakers of the language. Those phonetic features that distinguish phoneme categories are said to be distinctive. Systematic contextual variations in the phonetic realization of a phoneme are captured by allophonic rules; phonetic features that underlie only allophonic variations are said to be redundant. Finally, phonotactic and syllable structure rules specify constraints on the sequencing of phonetic segments in the language.

The languages of the world differ with respect to all three aspects of phonological structure: phoneme inventories, rule-governed allophonic variation, and phonotactic/syllable structure constraints. For any given language, the functional (phonemic) categories are a subset of those phonetic categories defined by the universal phonetic inventory. Thus, two phonetic segments that are distinctive in one language may not occur at all in another language, or may occur, but only as allophones (contextual variations) of a single phoneme. Finally, languages may differ in the syllable contexts in which particular phonetic segments can occur. The major focus of the research reported in this book is how the knowledge of all three aspects of the phonological structure of our native language affects the perception of spoken language.

When considering theoretical and empirical issues in the perception of speech, it is important to note that both phonetic and phonemic levels of analysis of speech are abstract. That is, linguistic utterances are represented as sequences of discrete, static segments, and articulatory features are used primarily as classificatory devices to describe linguistically relevant differences and similarities among phonetic segments. On the other hand, actual speech gestures are continuous and dynamic, and most importantly, temporally overlapping. That is, movements of the articulators associated with the realization of more than one phone of a phonetic sequence occur simultaneously (i.e., phones are coarticulated). Movements of multiple articulators are temporally coordinated, and they are usually characterized by smooth transitions between articulatory "postures" that are themselves timed. These rhythmic gestures give rise to the continuously varying pattern of acoustic energy that constitutes the speech signal.

Given the fact that phonetic sequences, defined either phonetically or phonemically, are abstract representations of actual speech utterances, it is perhaps not surprising that the acoustic signals generated by speech gestures cannot be analyzed in terms of an alphabet of discrete (and invariant) segments of sound that relate one-to-one to these abstract categories. However, the structure of the acoustic patterns is determined by (and can, therefore, be correlated with) properties of

the gestures, and in turn, to the abstract phonetic description of those gestures (Fant 1960). Thus, articulatory features that specify place-of-articulation of consonants and tongue position of vowels are related systematically to the spectral structure of the acoustic pattern (e.g., spectral peaks in release bursts, formant center frequencies); whereas, voicing and aspiration features of consonants can be related to sound source properties (periodic vs. aperiodic sound) and to temporal parameters (durations of silence or noise). Information about the manner of articulation is carried by both temporal parameters (e.g., formant transition durations, durations of noise) and by source characteristics. It must be noted, however, that these phonetically relevant acoustic properties are also characterized abstractly (relationally). For example, we can describe how the relative frequency of the first formant (F1) varies inversely with tongue height for vowels, but the exact frequency of F1 of the "same" vowel differs significantly as a function of the phonetic context in which the vowel occurs, with the rate of speech, and with the identity of the speaker.

In exploring the constancy problem in speech perception then, we must continue to address questions about whether the descriptions of the objects of perception and the descriptions of the acoustic signals as stimuli are at the appropriate levels of abstraction. Early studies of the acoustic properties of speech showed that the "acoustic cues" by which phonetic segments could be perceptually differentiated varied significantly as a function of the position of the segment in the syllable. Most perceptual studies, therefore, utilized stimuli in which the phonetic segments of interest were presented in a particular syllable context; for example, stop consonants in consonant-vowel (CV) syllables. Thus, although it is often tacitly assumed that a phonemic level of analysis is appropriate for characterizing the objects of speech perception, in fact, almost all empirical studies have been pursued at the level of the position-dependent allophone (c.f. Flege, this volume; Rochet, this volume).

In cross-language studies of speech perception, it is even more obvious that a phonemic level of analysis is too abstract to capture many of the phenomena of interest. Contrastive analyses of the phoneme inventories of different languages do not contain enough detail about the articulatory or acoustic structure of phonetic segments to allow researchers to make informed predictions about perceptual patterns or possible learning difficulties. This can be illustrated by the well-known example of "liquid" consonants in Japanese and American English.

The phoneme inventory of Japanese (J) includes a single liquid phoneme, usually transcribed as /r/. In contrast, American English (AE) is said to have a phonemic contrast between two liquids, /r/ versus /l/. However, this analysis misses several significant facts about the two phonologies. First, J /r/ is phonetically realized most often as an alveolar (retroflex) tap [ɾ] rather than as a postalveolar retroflex (or tongue bunched) approximant as in AE [ɹ] (i.e., J and AE /r/ differ in both place and manner of articulation). J /r/ sometimes includes a lateral release, making it similar in tongue shape to AE /l/. AE /l/ has two major allophones: alveolar lateral "light" [l] in syllable-initial (prestressed) position and velarized "dark" [ɫ] in syllable-final (poststressed) position. Acoustically, AE syllable-initial [ɹ] and [l] differ systematically in temporal as well as spectral properties of formant transitions (Dalston 1975); whereas, syllable-final [ɹ] and [ɫ] differ primarily in spectral structure. In addition, there is considerable anticipatory coarticulation of both syllable-final liquids, so the acoustic structure of the preceding vowels is strongly influenced by the upcoming consonant.

On the basis of the phonemic contrastive analysis, we might predict that Japanese learners of English will only have difficulty learning to produce English /l/ (but not /r/) and difficulty learning to perceive the English /r–l/ contrast in all syllable contexts. However, the more detailed phonetic contrastive analysis would yield predictions that more closely resemble the facts: 1) Japanese have difficulty producing both [ɹ] and [l]; 2) perceptual difficulties are significantly greater for syllable-initial than syllable-final liquids; and finally, 3) Japanese have most difficulty perceptually differentiating [ɹ–l] in prestressed consonant clusters—syllable structures that are not phonologically admissible in Japanese.

To summarize, important questions concerning the appropriate units/levels of analysis continue to be debated in current theories of perception of spoken language. These include how best to represent the objects and stimuli of perception and how linguistic representations of phonetic categories and sequences relate to spoken utterances as actually produced and perceived. Cross-language studies of speech perception bring a special perspective to these questions. Several alternative views of how linguistic (phonetic/phonological), gestural, and acoustic spheres of analysis interrelate are presented in subsequent chapters of this book. As researchers attempt to explicate the role of linguistic experience in the development and modification of speech perception, answers to questions about what is learned and how it is learned will shed light on these very basic issues.

THEORIES AND METHODS IN EARLY CROSS-LANGUAGE RESEARCH

Early interest in cross-language studies of speech perception stemmed in part from the theoretical claims of the motor theory of speech percep-

tion (c.f. Liberman et al. 1967). According to this theory, speech perception was special in that the processes by which the linguistic message was recovered from acoustic signals were hypothesized to be different from auditory processes used to perceive nonspeech acoustic signals. Specifically, early versions of motor theory postulated that context-conditioned (variable) speech sounds were perceived with reference to the production processes by which the phonetic segments were "encoded" in the first place. That is, perception of speech sounds was mediated by knowledge of how those sounds were produced by the articulatory system.

These claims were founded on empirical investigations of the perception of synthetic speech stimuli in which phonetically relevant acoustic parameters were systematically varied, and perception was assessed. Research on stop consonants in English revealed that perceptual categorization of a series of synthetic stimuli that varied along an acoustic continuum showed marked discontinuities. For instance, stimuli that differed only in the onset, direction, and extent of second formant (F2) transitions from low onset, rising transitions (typical of labial stops), through midfrequency onsets with slightly rising or falling formants (characteristic of alveolar stops) to high onset, rapidly falling transitions (underlying velar stops) were presented to subjects. The stimuli differed in equal frequency steps of F2 onset frequency, forming an acoustic continuum that encompassed three phonetic categories. However, listeners did not hear a continuously varying set of stimuli. Rather, they heard a series of indistinguishable /b/s, followed by a set of /d/s, then a set of /g/s. In other words, perceptual discontinuities seemed to correlate with the distinctive nature of the articulatory gestures that produce the different patterns of F2 transitions in natural speech.

This perceptual phenomenon was empirically verified by an experimental method that came to be known as the categorical perception (CP) paradigm. In this paradigm, listeners are tested on identification and discrimination of a set of synthetic stimuli that vary in physically equal steps along one or more acoustic dimensions underlying a phonetic contrast. In the identification test, listeners are asked to label the stimuli, presented one at a time in random order, using phonetic labels provided by the experimenter (i.e., using a forced-choice response format). In the discrimination task, two or more stimuli are presented sequentially within a trial, and listeners are asked to make comparative judgments about the physical identity or difference of the stimuli. The stimuli being compared in each trial differ by the same amount, usually stated in terms of the number of steps (sequential stimuli) by which each comparison pair differs (e.g. one-step pairs are adjacent stimuli; two-step pairs are stimuli 1–3, 2–4 ... 8–10). All

possible pairs of stimuli differing by that amount are tested repeatedly, and comparison pairs from different parts of the continuum are presented in random order.

Performance on the two tests is then compared. Typical results for voicing and place-of-articulation contrasts among highly encoded (context-dependent) speech sounds such as stop consonants are shown in Figure 1. Identification functions are marked by abrupt

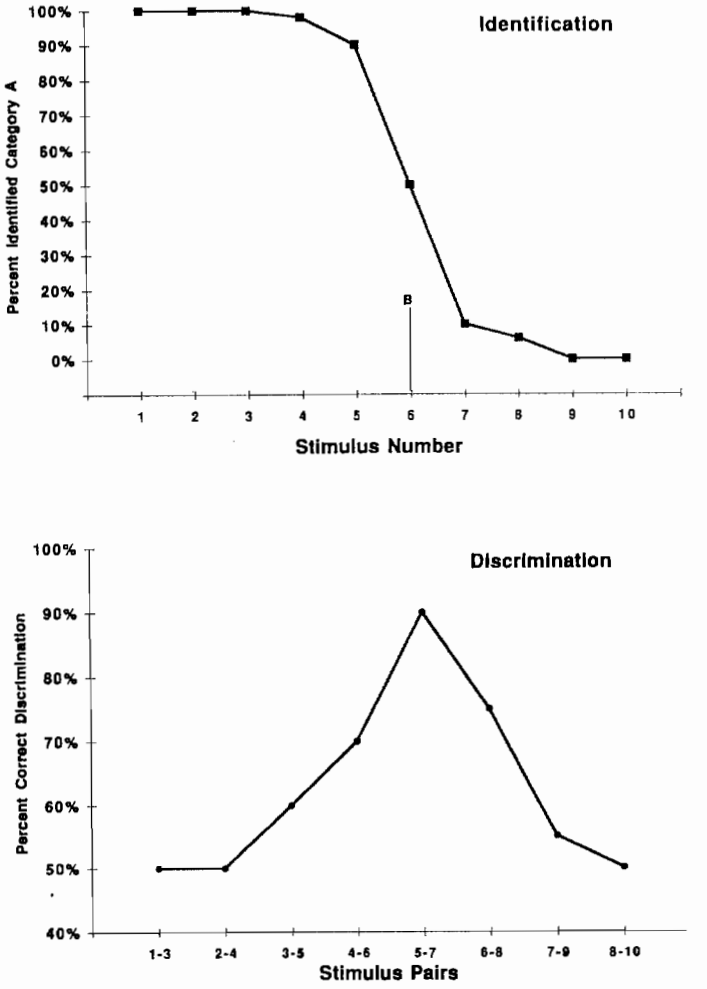

Figure 1. Categorical perception of consonant contrasts. Identification function (above) shows phoneme boundary (50% crossover) at #6; relative discrimination function (below) shows poor within-category discrimination.

boundaries between categories with highly consistent labeling of all stimuli within each category. Discrimination functions show correlated peaks of accurate performance for comparison pairs whose members are labeled as different phonetic segments (cross-category comparisons) and troughs of very poor discrimination of comparison pairs whose members are labeled as the same phonetic segment (within-category comparisons). If the location of peaks and troughs in discrimination functions can be predicted from labeling performance alone, the acoustic continuum is said to be perceived categorically.[1]

Many acoustic dimensions contrasting consonants have been shown to be perceived categorically, including F3 transition cues for the [ɹ-l] contrast (Miyawaki et al. 1975), F2/F3 transition cues for place contrasts among both oral and nasal stops (Mattingly et al. 1971; Miller and Eimas 1977; Pisoni 1973), transition duration cues for the stop versus approximant manner contrast (Miller and Liberman 1979), and voice-onset time (VOT) cues for voicing distinctions in stop consonants in syllable-initial position (Abramson and Lisker 1970).

In contrast, acoustic continua underlying other phonetic contrasts are perceived continuously, as shown in Figure 2. For instance, a F1/F2 frequency continuum underlying contrasts among isolated (uncoarticulated) steady-state vowels yields very different results (Fry et al. 1962; Pisoni 1973). First, identification of stimuli near the phonetic boundary is typically less consistent, resulting in more gradually sloping identification functions (A). More importantly, discrimination is not predictable from identification performance. For long duration steady-state vowels, discrimination of both within-category and cross-category comparisons is quite good (B). That is, discrimination is much better for within-category pairs than would be predicted on the basis of labeling data. This pattern of continuous perception has also been demonstrated for (preceding) vowel duration cues for the voicing contrast in final stop consonants of English (Raphael 1972).

[1]According to the strong version of the categorical perception hypothesis, discrimination performance is predictable on the basis of identification performance alone. Predicted functions are generated using the probabilities of labeling each member of a comparison pair as the same phoneme. If predicted and obtained functions do not differ significantly, the strong form of the CP hypothesis is accepted. Typically, obtained and predicted functions are very similar in overall shape; however, obtained functions often reveal somewhat better discrimination than predicted. This suggests that subjects are able to discriminate physical differences that are not phonetically relevant. A weaker form of the CP hypothesis states that if the peaks and troughs of discrimination functions can be predicted from identification, and if discrimination is signficiantly better for cross-boundary comparisons than for within-category comparisons, the continuum is perceived categorically.

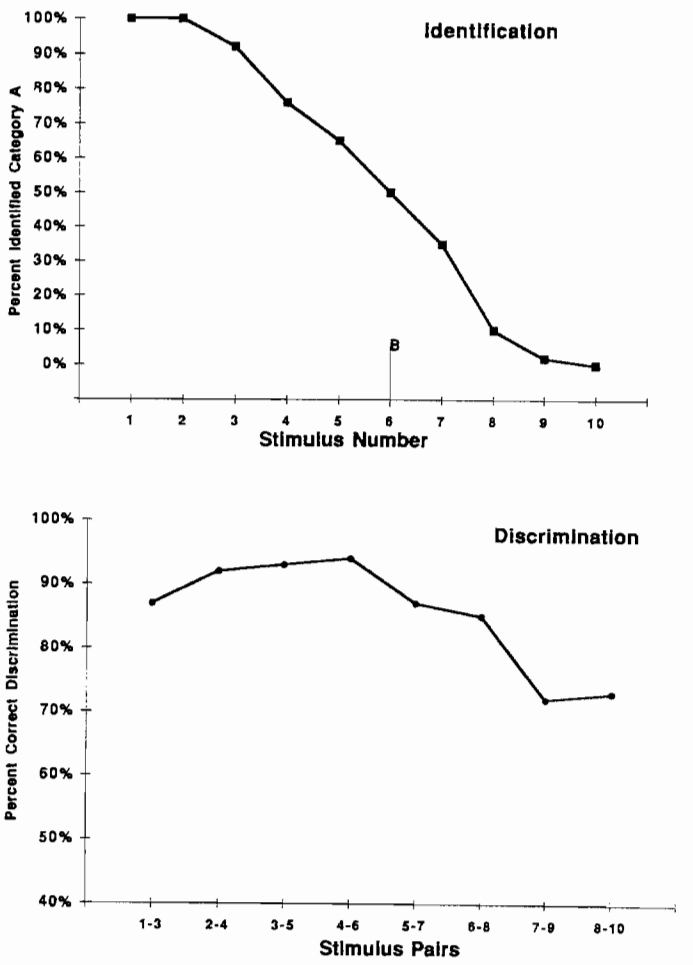

Figure 2. Continuous perception of (steady-state) vowel contrasts. Identification function (above) shows phoneme boundary (50% crossover) at #6; relative discrimination function (below) shows good within-category discrimination.

The continuous pattern of discrimination of formant frequency and duration parameters of vowels is much more typical of the perception of acoustic dimensions that distinguish nonspeech sounds. That is, the ability to discriminate physical differences between sequentially presented acoustic stimuli is usually much better than the ability to make absolute judgments, as required in a labeling task. In fact, this is true for some phonetically relevant acoustic components of

speech stimuli (such as F2 transitions) when they are extracted from the full speech patterns and presented as nonspeech "chirps" (Mattingly et al. 1971). Discrimination of the isolated components typically shows a continuous pattern of discrimination across the acoustic continuum (i.e., either uniformly high performance, uniformly low performance, or a monotonically increasing or decreasing function).

The categorical perception of acoustic continua underlying contrasts among encoded speech sounds was taken as evidence of the link between perception and production of speech. It also supported the notion of a specialized "mode" of perception that was engaged only when stimuli were sufficiently speech-like; that is, heard as sounds that had been produced by human articulatory gestures.

Much research in the 1970s was dedicated to questions about whether categorical perception was indeed unique to speech sounds and unique to humans' perception of speech, as implied by motor theory. (See Strange and Jenkins 1978, Repp 1984 for extensive reviews of this research.) To summarize very briefly, research on the perception of such nonspeech analogs as tone-onset time analogs of VOT (Pisoni 1977) and of musical chords (Burns and Ward 1973, 1975) demonstrated that categorical-like discrimination of nonspeech acoustic dimensions could be obtained. Research on the perception of synthetic speech continua by nonhuman mammals also revealed discontinuities in discrimination or "identification" performance strikingly similar to those shown by human adults (e.g., Kuhl and Miller 1978).

These findings led to an alternative account of the categorical perception of speech according to which perceptual discontinuities were attributed to the presence of "natural" boundaries in acoustic dimensions that distinguished complex acoustic patterns. This partitioning of the auditory perceptual space into natural categories was hypothesized to be a function of the biological preprogramming of mammalian auditory systems. According to this theory, the phonemic inventories of languages have evolved to take advantage of these innately given perceptual categories (c.f., Miller et al. 1976).

Given these conflicting explanations of the CP phenomenon, it was important to examine the discrimination of phonetically relevant acoustic continua across different language groups and across different age groups of humans. From the perspective of the early motor theory account, it was expected that adults who spoke different languages might show language-specific patterns of perception, depending on the linguistic function of the phonetic segments in their language. Furthermore, if perception of encoded phonetic segments depends on knowledge of their production, prelinguistic infants might not show categorical-like patterns of discrimination. If, on the other hand, CP reflects (language-independent) auditory perceptual

categories, both infants and adults might be expected to show similar patterns of perception, regardless of their particular language environment and experience. Cross-language studies of speech perception by adults and infants, thus, provided important empirical evidence that shaped the theoretical debates of the day.

FINDINGS AND CONCLUSIONS OF EARLY CROSS-LANGUAGE RESEARCH

With some notable exceptions (Briere 1966; Goto 1971; Trehub 1976), the CP paradigm was the dominant method used in cross-language studies of perception in the 1960s and 1970s. Typically, two language groups were tested on the same set of synthetic stimuli, which varied along a phonetically relevant acoustic continuum (see Beddor and Gottfried this volume). The groups differed in the linguistic function of the phonetic categories spanned by the continuum (i.e., whether the phonetic segments were distinctive, allophonic variations, or did not occur at all in the listeners' native language). It was assumed that all subjects were equivalent with respect to their general auditory perceptual capabilities. Thus, any differences across groups in discrimination of the stimuli could be attributed to the influence of language-specific experience on phonetic perceptual processes.

Developmental studies with infants were also conducted using the same synthetic stimuli. Patterns of relative discrimination of pairs of stimuli that, for adult listeners, constituted cross-category versus within-category comparisons were examined to determine whether and when prelinguistic infants showed perceptual discontinuities similar to adults. A few cross-language infant studies were also undertaken, in which infants from a language environment that did not utilize the phonetic contrast under study were compared with those from a language environment in which the contrast was phonemic.

The CP paradigm was also employed to investigate changes in perception by learners of a second language (L2), and to explore how perceptual training affected the perception of synthetic continua underlying non-native phonetic contrasts. Results of these early cross-language studies with adults, infants, and L2 learners are summarized briefly, and conclusions drawn from these studies are presented.

Cross-Language Differences in Adult Perception

Many of the early cross-language studies examined the perception of the VOT synthetic speech continua developed by Lee Lisker and Arthur Abramson. Acoustical analyses of productions of stop consonants in 11 languages by these researchers demonstrated that differences in VOT were sufficient to distinguish both voicing and aspiration contrasts among syllable-initial oral stops in most of the languages (Lisker

and Abramson 1964). To conduct perceptual studies, Abramson and Lisker synthesized three series of C + /ɑ/ syllables in which the onset of voiced (periodic) energy in the frequency region of F1 was varied relative to onset of the release burst and upper formant energy (F2/F3). A labial series, an alveolar series, and a velar series were generated by adjusting the F2 and F3 transitions appropriately. Each series ranged from an extremely prevoiced stimulus in which voicing murmur preceded the release burst by 150 ms (150 ms of voicing lead or –150 VOT) through simultaneous onset of F1 and upper formants immediately following the release burst (0 VOT) and then continuing to an extremely postvoiced stimulus in which the onset of F1 and periodic source for upper formants occurred 150 ms after the release burst (150 ms of voicing lag or +150 VOT). Each stimulus differed from adjacent stimuli in duration of voicing lead (prevoicing) or voicing lag (postvoicing) in 5 or 10 ms steps. The synthetic series encompassed so-called fully voiced (prevoiced) stops, devoiced or voiceless unaspirated (simultaneous or short voicing lag) stops, and voiceless aspirated (long voicing lag) stops.[2]

Identification tests using these stimuli (Abramson and Lisker 1970, Williams 1977) showed that adult listeners divide the VOT continuum into two or three categories, depending on their native language. English speakers categorize the stimuli as either voiced or voiceless, with a boundary between categories at +20 to +40 VOT, depending on the place of articulation. That is, both prevoiced and short lag stimuli are heard as voiced /b, d, g/; whereas, stimuli with long voicing lags are labeled as voiceless /p, t, k/. In contrast, Spanish speakers label stimuli with long voicing leads as /b, d, g/; whereas, stimuli with very short voicing leads or voicing lags of any length are labeled as voiceless /p, t, k/. The boundary between categories falls at about -5 VOT. Finally, speakers of Thai, for whom both voicing and aspiration contrasts are phonologically distinctive in bilabial and alveolar stops, identify three distinct categories (voiced, voiceless unaspirated, voiceless aspirated), with boundaries at about -20 VOT and + 40 VOT, respectively. (The velar series was divided into two categories by Thai speakers, following the phonological rules of their language.)

Discrimination tests of the VOT continuum by speakers of these languages reveal that VOT is perceived categorically. Peaks of accurate discrimination occur only for comparison pairs that are drawn from opposite sides of a phonetic boundary that is distinctive in the

[2]Note that on the prevoiced side of the continuum, stimuli differed only in duration of the voicing murmur preceding the release. On the postvoiced side, stimuli differed in the amount of F1 cutback and the duration of aperiodic sound (aspiration) between the transient release burst and beginning of the periodic sound source.

listeners' native language. Thus, English and Spanish speakers each show a single discrimination peak; whereas, Thai speakers show two peaks for labial and alveolar stops. The location of the peak(s) for all three language groups is predictable from identification functions. Discrimination of stimuli that are not distinctive in the listener's native language is relatively poor.

In the case of VOT, then, both the presence and location of discontinuities in discrimination functions are determined by language-specific (phonological) experience. An early cross-language study of the place-of-articulation distinction between AE liquids /ɹ-l/ by native speakers of AE and Japanese illustrates a similar language-specific pattern of perception (Miyawaki et al. 1975). As described in a preceding section, this distinction is not phonemic in Japanese, nor do either of these phonetic segments occur in that language. Discrimination tests, using a synthetic [ɹɑ-lɑ] series in which only the F3 transitions varied, revealed significant differences between the two language groups. American English listeners showed a peak of accurate discrimination for cross-category comparison pairs and troughs of less accurate discrimination of within-category pairs (i.e., they perceived the continuum categorically). In contrast, Japanese listeners showed quite poor discrimination of all comparison pairs, resulting in large differences between groups in discrimination of cross-(AE)-category comparison pairs. In addition, tests of discrimination of the F3 components taken out of speech contexts (where they sounded like nonspeech glissandi) revealed that both Japanese and AE listeners could discriminate F3 transition differences across the entire continuum with fairly high accuracy. Thus, differences in discrimination of the speech stimuli by the two language groups reflected differences in phonetic, rather than auditory processing of the F3 transition cue.

An early cross-language study of the discrimination of synthetic (steady-state) vowel continua by native speakers of AE and Swedish revealed a different pattern of perception (Stevens et al. 1969). Two series of vowels were generated that each varied in F1/F2/F3 frequency over a range that encompassed three vowel categories. One series contrasted three front unrounded vowels that are phonologically distinct in both languages; the other series included front rounded vowels that are phonemic in Swedish, but not in AE. In contrast to the cross-language studies of consonant contrasts, discrimination of both vowel series by AE and Swedish listeners did not reveal language-specific effects. Both groups showed continuous (and quite accurate) discrimination along the entire range of both continua. There were no significant differences in performance on cross-category versus within-category pairs for either group and no differences between Swedish and AE subjects in discrimination of the front rounded vowel stimuli.

From these seminal cross-language CP studies, it was concluded that knowledge of the native language phonological system influenced adults' ability to discriminate some, but not all, phonetically relevant acoustic parameters. The presence and location of discontinuities in the perception of acoustic dimensions contrasting consonants were predictable from the linguistic function of the phonetic contrasts. On the other hand, synthetic vowel continua did not show the same effects of language-specific experience. This gave credence to the hypothesis that language experience shaped the special perceptual processes used to decode the rapidly varying and highly context-dependent acoustic parameters that distinguished consonants. The perception of steady-state vowels (and nonspeech sounds), which did not require special decoding mechanisms, appeared not to be influenced by linguistic experience in the same way.

Discrimination of Native and Non-Native Contrasts by Infants

The results of cross-language studies with adult monolinguals might have led to the conclusion that the language-specific perceptual discontinuities associated with consonant contrasts were a function of learning one's native language were it not for the results of infant studies being conducted at the same time with some of the same stimuli. In short, studies using both VOT continua and the [ɹɑ-lɑ] continuum showed that AE infants as young as 1 to 2 months of age showed categorical-like patterns of discrimination of these continua. For instance, using the high-amplitude sucking habituation paradigm (c.f. Polka et al., this volume) Peter Eimas and his colleagues (Eimas et al. 1971) demonstrated that very young infants from English-speaking environments discriminated labial stops that differed in VOT by 20 ms only when the stimuli constituted a cross-English-category /b–p/ comparison (+20/+40 VOT). Discrimination by infants tested on a within-/p/ pair (+60/+80 VOT) or a within-/b/ pair (-20/0 VOT) was poorer and did not differ from infants in the no-change control group.

This study was extended to cross-language comparisons of infants from language environments in which short lag versus long lag stimuli did not constitute a phonemic contrast. Lasky and his colleagues (Lasky, Syrdal-Lasky, and Klein 1975) demonstrated that 6-month-old infants from a Spanish-speaking environment discriminated the English contrast in labial stops, but not the Spanish contrast. Streeter (1976) also reported that infants from a Kikuyu-speaking environment could discriminate the short-lag/long-lag labial stops. This is quite interesting, because Kikuyu uses only a single (prevoiced) labial stop. These studies demonstrated that prelinguistic infants could discriminate VOT differences between voiceless unaspirated (short lag) and aspirated (long voicing lag) stops, whether or not they had been exposed to both types of phonetic segments in their language environment.

In one of the few early studies that used natural speech tokens, Trehub (1976) investigated Canadian English-learning infants' perception of two non-native contrasts: French oral versus nasal vowels /papã/ and Czech palatal fricative versus fricative vibrant (trill) /ʒ-r̆/. Both contrasts were discriminated by 2- to 4-month-old infants; whereas, adult Canadian English listeners could not discriminate the Czech contrast, and their performance on the French vowel contrast, although above chance, was poorer than that of the infants.

On the basis of these cross-language studies, as well as additional studies of English-learning infants on English contrasts, it was concluded that infants could perceptually differentiate almost all of the phonetic contrasts of the adult language at a very early age. Furthermore, it was concluded that speech perception by prelinguistic infants was adult-like, because discrimination patterns showed discontinuities in the perception of acoustic continua underlying consonant contrasts (i.e., discrimination was categorical). Finally, cross-language studies suggested that infants were "language-universal" perceivers; phonetic contrasts were perceptually differentiated, regardless of their phonological status or even their occurrence in the adult language to which the infants had been exposed. At this early age, perception was not yet affected by specific linguistic experiences, but rather, reflected innate language-learning abilities.

Between early infancy and adulthood, then, children's interactions with their linguistic environment while acquiring their first language produce significant changes in the perception of speech sounds. There is a "loss" in the ability to differentiate phonetic categories perceptually that are not phonologically distinctive in the native language, while native contrasts may become more highly differentiated. Given this pattern of results, several questions about the nature and timing of these developmental changes were pursued. What was the nature of the loss of discrimination? When did the shift from language-universal to language-specific perception take place? How was this developmental pattern related to other aspects of language learning, including production of speech sounds and lexical learning? Research exploring these questions is described in a later section of this chapter; however, first, early perceptual research on second-language learners is summarized.

Perception of Non-Native Contrasts by Second-Language Learners

It is well known that adult L2 learners have difficulty learning to produce some non-native phonetic segments, which leads to the persistence of accented pronunciation in the L2. Given the results of

cross-language perception studies, it was also reasonable to predict that adult L2 learners might also have difficulty learning to perceive some non-native contrasts. An early study by Goto (1971) documented the startling fact that Japanese learners of English who had learned to produce AE [ɹ] and [l] distinctively (as judged by native English listeners), nevertheless, still had difficulty perceptually differentiating the liquids in recordings of their own speech and the speech of native AE speakers. In other words, in the case of the [ɹ–l] contrast, at least, it appeared that some native Japanese speakers had learned to produce this difficult contrast before they had learned to perceive it by auditory means. (See Sheldon and Strange 1982 for a replication of this result.)

Early CP studies documenting changes in the perception of non-native phonetic contrasts by L2 learners were, again, dominated by studies of VOT. Williams (1979) reported a cross-sectional study of native Spanish-speaking children (ages 8–10 years and 14–16 years) learning English in the United States. Perceptual boundaries between voiced and voiceless labial stops showed a gradual shift away from the Spanish boundary (-4 VOT) toward the native English boundary (+25 VOT) as a function of the time spent in the United States. However, even after 3 years of English experience, identification functions were not the same as those of monolingual English speakers (phonetic boundaries averaged +12 VOT for younger children, +9 VOT for the older children). Streeter and Landauer (1976) also reported gradual improvement in the perception of the English contrast between labial stops by Kikuyu children learning English in school in their native country.

These studies suggested that the modification of native-language perceptual patterns was possible, at least for children, but occurred gradually over the course of learning a second language. Williams' data also suggested that preadolescent children's perception of non-native contrasts improved more rapidly than did older children's, supporting the theory that there was a critical or sensitive period for language learning (Lenneberg 1967). Questions about the nature and time course of changes in phonetic perception during L2 learning and about the relationship between perceptual change and production of non-native contrasts have been pursued more recently; however, before discussing this research, one more strand of early cross-language research must be summarized.

Perceptual Training of Non-Native Phonetic Contrasts

Early theoretical arguments about the nature of the CP phenomenon led to experimentation on the malleability of language-specific patterns of discrimination demonstrated with synthetic stimuli series such as VOT (c.f. Lane 1965 versus Studdert-Kennedy et al. 1970).

Although not directly motivated by questions about perceptual change during L2 learning, these early studies contributed to the thinking of the time about the nature of the language-specific patterns of perception demonstrated by adults. Once again, studies of VOT provided the earliest evidence about the malleability of phonetic perceptual patterns. Strange (1972) investigated whether AE speakers could be trained to discriminate differences in VOT that constituted contrasts in Thai and Spanish, but not in English, using the Abramson and Lisker stimuli and several kinds of perceptual tasks. She reported that although some improvement in discrimination or identification of short lag versus short lead stimuli occurred, there was very little evidence of transfer of training to any stimuli other than those used during training. Lisker (1970) had earlier reported a failure to train Russian listeners to differentiate the short lag versus long lag (English) contrast. This result was especially interesting because studies of human infants and of nonhuman mammals had shown good discrimination of this difference in VOT.

From these very limited data, it was concluded that changing phonetic perceptual patterns in adults by intensive short-term training procedures was very difficult, if not impossible (however, see Carney, Widin, and Viemeister 1977). Improvement in performance appeared to be restricted to the training stimuli and, in some cases, even to the specific perceptual tasks used in training; that is, there was no generalization to novel stimuli or situations. This suggested that subjects had not altered their phonetic perceptual patterns, but rather had learned only to attend to the specific aspects of the training stimuli. The relevance of this type of perceptual change to the broader question of how perception changes with second-language learning, thus, appeared to be rather limited.

Conclusions and Limitations of Early Cross-Language Research

Cross-language studies of VOT and a few other phonetically relevant acoustic dimensions revealed language-specific patterns of perception of consonantal contrasts by adult listeners. In contrast, noncategorically perceived steady-state vowel continua and nonspeech analogs did not reveal language-specific perceptual patterns. Thus, studies of prelinguistic infants and L2 learners concentrated almost exclusively on patterns of perception of consonants, especially voicing contrasts among stops.

The dramatic findings of discrimination studies with very young infants led to the conclusion that humans beings come into the world equipped to differentiate perceptually many, if not all, the phonetic categories that can function to distinguish lexical items in any language. The developmental process, then, was conceived of primarily

as one of selective "loss" of the ability to differentiate those contrasts that were not functional in the learner's native language. Studies of L2 learners and training studies with monolinguals suggested strongly that these well-learned language-specific patterns of perception were highly resistant to modification in adulthood. Thus, adult L2 learners were characterized as having "accented" perception as well as accented production.

These conclusions were based on very limited data exploring only a handful of phonetic contrasts. The dominant methodology was the CP paradigm which employed synthetic stimuli. There was very little research in the 1960s to 1970s that utilized natural stimulus materials or perceptual tasks that might have tapped perceptual processes at different levels. In retrospect, it is clear that several of the conclusions were overstated and premature. However, the continuing general theoretical concern about the interaction of "nature" and "nurture" in phonetic perception led to expanded research efforts. New contrasts among both vowels and consonants were investigated, using natural speech stimuli, as well as synthetic speech continua where multiple acoustic parameters were manipulated orthogonally. New tasks were utilized to tap into perceptual processes at different levels. Cross-language studies with new language groups and listeners of many ages explored questions about the nature and malleability of the language-specific patterns of perception. In the next section, some of the most influential findings of research in the 1980s and early 1990s are reviewed.

FINDINGS AND CONCLUSIONS OF RECENT CROSS-LANGUAGE RESEARCH

In the 1980s and early 1990s, theoretical debates about the nature of the mechanisms involved in the perception of speech continued to be articulated from the perspective of three major points of view: the (revised) motor theoretic account, the "psychoacoustic" or "general auditory" account, and the direct realist account. (See Best, this volume, for a comparison of these three viewpoints as they relate to cross-language speech perception research). Paradigms developed to investigate models of the relative contribution of "auditory" and "phonetic" levels of processing of speech were also used to explore cross-language differences in phonetic perception. In addition, researchers primarily interested in theoretical and empirical questions concerning the influence of linguistic experience on speech perception expanded their investigation to new types of phonetic contrasts and comparisons of new language groups. In this section, some of the methodological developments that contributed to our increased understanding of the phenomena of cross-language perception are described. (See the chapters in this volume by Polka et al., Beddor and Gottfried, and Logan and Pruitt for more detailed descriptions and critiques of current methods in cross-language studies of speech perception.) Following this, some of the major findings of recent research on cross-language perception in adult monolinguals, infants, and L2 learners are reviewed and conclusions that have been drawn from this research are presented.

Methodological Developments in Cross-Language Research

Synthetic speech stimuli continued to be exploited in speech perception research in the 1980s and 1990s. The use of synthetic stimuli allows for a precise description of the acoustic basis for perceptual differentiation. In addition, acoustic synthesis allows investigators to manipulate independently the individual acoustic parameters that are typically coupled or correlated in speech produced by a talker. One method that exploited this advantage was the "trading relations" paradigm. As an example, Polka and Strange (1984) generated two synthetic series contrasting AE [ɹ-l]; in each series the F3 (and F2) onset and transition (spectral) cues varied concurrently from [ɹ]-like to [l]-like values in 10 steps. The two series differed from each other in the duration of the F1 initial steady-state and transition. In the first series, the F1 temporal structure was patterned after an [ɹ] with a relatively short steady-state and slow transition into the following vowel. In the other series, the steady-state was long followed by a rapid transition, as is appropriate for a prevocalic [l] in AE.

Results of identification and discrimination tests of the 20 stimuli demonstrated that although the spectral differences provided the primary cues for differentiation of [ɹ-l] by AE listeners, the temporal cue also influenced perception. Stimuli with intermediate spectral values (i.e., stimuli close to the phonetic boundary on the spectral dimension) were identified as [ɹ] in the first series, but as [l] in the second series. Furthermore, discrimination tests in which spectral and temporal cues were paired in facilitating or conflicting combinations showed that the two acoustic cues were perceptually integrated. Figure 3 illustrates this result. When a stimulus with an [ɹ]-like spectrum and tempo was compared with a stimulus with [l]-like spectrum and tempo (facilitating cues), discrimination in the vicinity of the phonetic boundary was significantly better than when only the spectrum varied (one cue comparisons). In contrast, when a stimulus with an [ɹ]-like spectrum and an [l]-like tempo was compared with a stimulus with an [l]-like spectrum and an [ɹ] -like tempo (conflicting cues), discrimination was poorer than when only the spectrum varied. Such a pattern of results was taken as evidence that discrimination was not based on psychoacoustic differences (which were equivalent in facilitating and con-

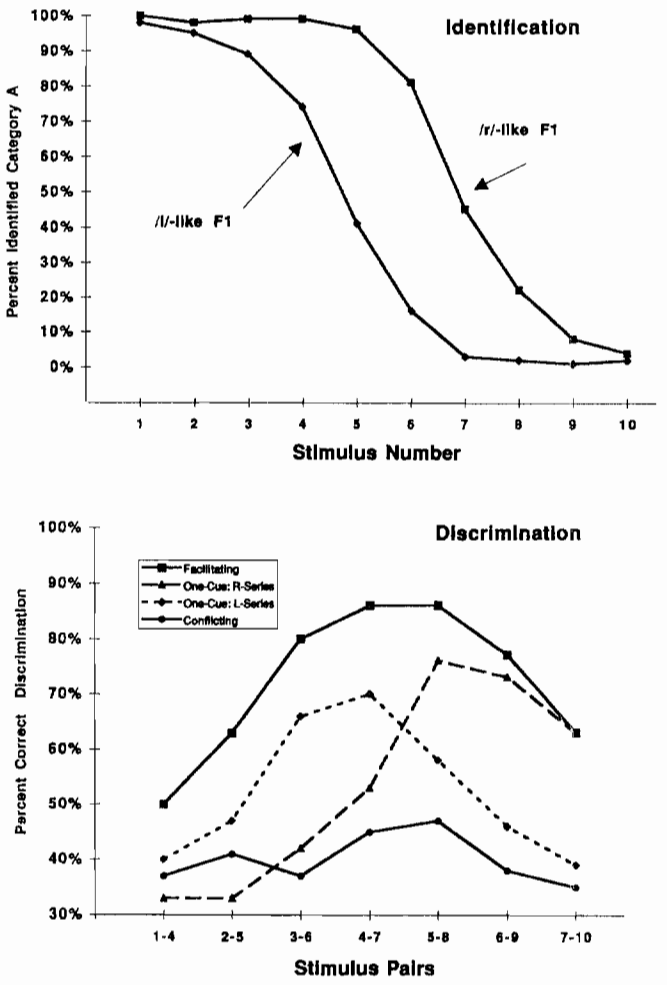

Figure 3. Trading relation between temporal and spectral cues for the /ɹ-l/ contrast. Identification functions (above) and discrimination functions (below). [Redrawn from Polka and Strange 1984.]

flicting cue pairs), but rather on the integrated phonetic percepts.

Although there was continued use of synthetically generated speech materials, cross-language studies with both infants and adults more often utilized carefully crafted sets of naturally produced stimuli. This allowed investigators to expand their investigations to phonetic categories/contrasts for which the relevant acoustic cues (for native speakers) were not well known. Stimulus materials typically included

several tokens of each phonetic category, spoken by one or more native talkers. The use of multiple talkers introduces physical variations in the phonetically relevant acoustic cues that require that listeners respond to relational rather than absolute acoustic parameters of the stimuli. The inclusion of multiple tokens by one or more talkers also introduces variations in phonetically irrelevant parameters of the stimuli, such as loudness, speaking rate, pitch, and intonation. Through careful analysis and selection of materials, these nonphonetic variations can be controlled or, more interestingly, varied independently of phonetically relevant parameters. In the latter case, the perceptual task becomes more ecologically valid, because listeners must respond on the basis of the (abstract) phonetically relevant information while ignoring irrelevant variations that normally occur (see Beddor and Gottfried this volume, for a more detailed description of these categorial tasks).

Several new types of perceptual tasks were developed to investigate both infant and adult listeners' ability to categorize these naturally varying speech utterances on the basis of their phonetic identity. Thus, subjects' ability to perform equivalence classification of discriminably different instances of phonetic categories was tapped. Studies of 6- to 12-month-old infants' categorization of multiple tokens of phonetic categories was examined using a variety of procedures (see Polka et al., this volume). *Categorial* (name-identity) identification and discrimination paradigms were used with adult monolinguals and L2 learners (see Beddor and Gottfried this volume). Finally, perceptual training studies employed categorization tasks rather than (physical-identity) discrimination tasks to investigate effects of short-term training on perception of non-native categories (see Logan and Pruitt this volume).

Cross-Language Studies of Adult Monolinguals

In the 1980s and 1990s, cross-language studies of an expanded set of phonetic contrasts revealed that the degree of difficulty adult listeners experienced in perceiving non-native consonants varied over a considerable range. Werker and Tees (1983, 1984) reported that English-speaking subjects failed to discriminate a velar–uvular place contrast in voiceless ejective stops [k'–q'] of Nthlakapmx (Salish) above chance levels. Polka (1992) found that AE speakers differentiated the same place contrast quite well (although still not as well as native speakers) when it occurred in Farsi voiced (egressive) stops [g–G]. Furthermore, Polka reported that Farsi speakers, for whom the place feature was distinctive for egressive stops, nevertheless did not perceive the same place contrast in Salish ejectives any better than did the AE speakers. Thus, it appears that the relative difficulty of a non-native contrast

cannot be predicted on the basis of whether or not the phonetic *feature* is distinctive in the native language. Rather, perceptual difficulty depends on the particular phonetic segments being contrasted and how they relate to the native language phoneme inventory.

In another set of studies, Janet Werker and her colleagues investigated the perception by English speakers of non-native voicing [d^h-t^h] and place [t̪–ʈ] contrasts in Hindi stop consonants (Werker and Tees 1983, 1984b; Werker et al. 1981). They found that although English listeners differentiated both contrasts more poorly than native Hindi speakers, the place contrast seemed to be more difficult than the voicing contrast. This may reflect the fact that the voicing contrast is similar to a phonemic distinction in English [d–t^h]; whereas, the place contrast distinguishes two phonetic segments that are similar to allophonic variations of a single English phoneme category (as in the words "width" [wɪd̪ð] versus "dry" [ɖɹaɪ]).

Polka (1991) replicated and extended Werker's investigation of the perception of Hindi dental versus retroflex stops by English speakers. She reported that categorial discrimination accuracy varied significantly as a function of the particular phonetic segments tested; subjects differentiated the contrast in voiceless unaspirated stops [t̪–ʈ] best (A′=.80); whereas, they performed no better than chance on the contrast in prevoiced stops [d̪–ɖ] (A′=.52). Performance levels for the contrast in breathy voiced [d̪ʱ–ɖʱ] and voiceless aspirated [t̪ʰ–ʈʰ] stops were intermediate. Pruitt (1992) also reported significant variation in the perceptual difficulty of Hindi dental versus retroflex stops as a function of the voicing/aspiration context, the following vowel context, and the particular Hindi talker producing the stimuli.

In all these studies of AE listeners' perception of the dental–retroflex contrast, performance by almost all subjects, even in the most favorable contexts, was significantly poorer than that of native speakers. In contrast, studies of other non-native consonant contrasts demonstrated that perception was sometimes highly accurate even when the phonetic segments were very unfamiliar to the listeners. Catherine Best and her colleagues (Best, McRoberts, and Sithole 1988) provided an important example of this pattern in a study of English speakers' categorial discrimination of nine place and nine voicing contrasts among Zulu (oral) clicks (a manner class characterized by interoral suction). In general, performance was very good, and for some contrasts, not significantly worse than that of native speakers. Performance on voicing versus place contrasts did not differ overall (mean = 92% correct for each type), although within each set of contrasts, some were more difficult than others (performance on individual contrasts ranged from 81% to 99% correct). Because all clicks were highly unfamiliar to English speakers *as speech sounds*, their relative difficulty

could not be predicted from phonetic similarity to native phoneme categories or to linguistic familiarity. However, some of the clicks were familiar as non-speech sounds (such as the "tsk, tsk" sounds made to express disapproval), and subjects reported that they heard all the stimuli as "clicks," "pops," "drips" and other non-speech sounds. They apparently employed these non-speech auditory properties in differentiating the categories.

Cross-language studies of vowels in the 1980s brought about a renewed interest in the perceptual difficulties posed by this class of speech sounds. Unlike an early study (Stevens et al. 1969) that showed no effect of native language on vowel perception, Beddor and Strange (1982) reported differences in the perception of the oral–nasal contrast in consonants [ba–ma] and vowels [ba–bã] by Hindi and English speakers. For both language groups, the consonant contrast is phonemic; whereas, the vowel contrast is phonemic for Hindi speakers but constitutes an allophonic variation for English speakers. Performance on a [ba–ma] synthetic series was equivalent for the language groups; both groups perceived the contrast categorically with only slight differences in the location of the category boundary. However, discrimination tests of the [ba–bã] series showed a different pattern. The contrast was perceived categorically by Hindi listeners, whereas, English listeners' discrimination functions were more continuous, showing better within-(Hindi)-category discrimination than Hindi speakers.

In a study that utilized natural speech materials, Gottfried (1984) also reported significant effects of native language on the perception of vowels. He compared categorial discrimination of eight contrasts among French vowels by English monolinguals and native French speakers. Both isolated vowels and vowels in /tVt/ syllables produced by several talkers were presented using a categorial ABX procdure in which the three stimuli of each trial were produced by different talkers, and listeners had to report whether the third vowel was an instance of the same phonetic category as the first or the second vowel. Even the native French-speaking subjects made perceptual errors on some of the vowel contrasts in this difficult task, especially when the vowels were in consonant-vowel-consonant (CVC) context. However, performance by the English monolinguals was, on average, significantly poorer. American English subjects had particular difficulty (relative to native speakers) with the contrasts between the front rounded vowels [y–ø] and between the front and back rounded vowels [y–u]. They also had difficulty differentiating the monophthongal [e] from [i] and from [ɛ].

This study demonstrated that when vowel categorization (as opposed to discrimination of physical differences) is examined, significant cross-language differences in perception are revealed. Some

vowels that do not occur phonemically in the native language are quite difficult for adults to categorize appropriately. In addition, vowels that do occur as phonemic categories in the native language, but differ in their phonetic detail between the native and non-native languages may also pose perceptual difficulties (see also the section below on L2 learners).

The above cross-language studies demonstrate that adult listeners have considerable difficulty perceiving many non-native contrasts among both vowels and consonants, especially when the stimulus materials incorporate the type of variability normally found in speech utterances (see Pisoni and Lively this volume). However, performance across a variety of non-native phonetic categories ranges from near native-like levels of accuracy to chance performance. Thus, the severity of the perceptual problem facing non-native listeners cannot be predicted merely from an analysis of the phoneme inventories of the native and non-native languages. Other factors that must be taken into consideration include phonetic similarities and differences between native and non-native phonemes, the allophonic distribution of phonetic segments in native and non-native languages, the influence of phonotactic factors (syllable position, phonetic context) on the articulatory and acoustic structure of phonetic segments, and dialectal/idiolectal variations that determine the intelligibility of individual native speakers (cf. Strange 1992). Finally, the "psychoacoustic salience" of the phonetically relevant acoustic cues may influence performance levels by non-native listeners (Best et al. 1988; Burnham 1986; Polka 1991). Current models that attempt to predict and account for the variations in the perceptual difficulty of non-native phonetic categories are presented in several chapters of this volume (cf. Best; Bohn; Flege).

The utilization (and, in some cases, comparison) of different types of speech perception tasks to assess cross-language perception also shed new light on the nature of the perceptual processes that are modified by experience with the native language. It appears that the perceptual difficulties of non-native listeners do not result from a loss in the sensory capacity to detect acoustic differences that are not used in contrasting phonemes in the native language. It can be demonstrated that under optimal listening conditions, adults can discriminate even the most difficult non-native contrasts with much the same accuracy as native listeners. For instance, Werker and Logan (1985) showed that English speakers could discriminate dental versus retroflex Hindi stops when the interstimulus interval was short enough to allow listeners to respond on the basis of "auditory" processing of the acoustic patterns. Thus, we can conclude that the ability to detect the phonetically relevant acoustic variations in speech utterances is not irretrievably lost in the course of learning our native language. Rather, language-specific

patterns of perception of phonetic contrasts reflect the attunement of selective perception (Strange 1986, 1992). The nature of these selective mechanisms and how they are modified by linguistic experience is discussed further below and also in several chapters of this volume (cf. Best; Kuhl and Iverson; Pisoni and Lively; Wode).

Cross-Language Studies of Infant Speech Perception

Some of the most important advances in the 1980s and 1990s in our understanding of the role of linguistic experience in the perception of speech stem from the results of cross-language studies of infants. Rebecca Eilers and her colleagues were among the first to report effects of language experience on the perception of consonantal contrasts by infants under 1 year of age. For instance, in a comparison of infants from monolingual Spanish versus monolingual English homes, Eilers, Gavin, and Oller (1981) reported that 6- to 8-month-old Spanish-learning infants could discriminate the Spanish tapped /r/ [ɾ] versus the trilled /r/ [r] as well as the Czech [ʒ–ř] contrast and the English [s–z] contrast. In contrast, the English-learning infants discriminated the Czech and English contrasts, but not the Spanish contrast. Eilers, Gavin, and Wilson (1979) also reported that Spanish-learning 6- to 8-month olds discriminated the Spanish voicing contrast (-20/+10 VOT), while English-learning infants the same age did not. (Both groups discriminated the English contrast [+10/+40 VOT].) These studies suggested that even in the first year of life, infants' perception was being modified by experience with the native language. Although some non-native contrasts were still differentiated, other contrasts were differentiated only if they occurred in the child's language environment.

Janet Werker and her colleagues performed both cross-sectional and longitudinal studies that examined the time course of the change in phonetic perception from a language-universal pattern to the language-specific patterns produced by adults. Her studies of English-learning infants' perception of Salish and Hindi place contrasts revealed that while 6- to 8-month-old English-learning infants could differentiate both these non-native contrasts, 11- to 13-month-old infants no longer differentiated either non-native place contrast (Werker and Lalonde 1988; Werker et al. 1981; Werker and Tees 1984). The English-learning infants continued to differentiate a native English place contrast [b–d], and Hindi-learning and Salish-learning infants continued to differentiate their native place contrasts. Thus, the decline in phonetic perception shown by 1-year olds was not attributable to a general decline in "auditory attention," but rather reflected the development of selective patterns of perception.

This interpretation was reinforced in a study by Best and her colleagues (Best et al. 1988) on English-learning infants' perception of Zulu

clicks. Just as for the adult English speakers, even the most difficult place contrast (voiceless unaspirated apical vs. lateral clicks) was discriminated by infants from 6 through 14 months old. We can assume that both infants and adults were able to respond accurately on the basis of salient psychoacoustic dimensions of the stimuli.

On the basis of these studies of consonantal contrasts, it was concluded that language-specific patterns of phonetic perception begin to emerge in the second half of the first year of life. Many non-native place and voicing contrasts that can be discriminated by 6-month olds, are no longer differentiated by 12-month olds. In addition, the location of "natural" phonetic boundaries may undergo shifts in the first year of life as a function of specific linguistic experience.

More recently, developmental cross-language studies of vowels have been pursued and reflect a different developmental time course (see Kuhl and Iverson this volume; Werker this volume; Werker and Polka 1993). For instance, Polka and Werker (1994) reported that English-learning 4-month-old infants discriminated non-native contrasts in German vowels better than 6-month olds, with a further decline in perception by 12-month olds. These results suggest that language-specific patterns of selective perception begin to emerge earlier for vowels than for consonants. Patricia Kuhl's finding of a language-specific "magnet effect" in the internal organization of vowel categories by 6-month olds corroborates this conclusion (Kuhl et al. 1992; Kuhl and Iverson this volume).

Peter Jusczyk and his colleagues have explored the developmental course of perception of more global properties of speech such as stress patterns, syntactic juncture, and intonational contours. He has also looked at the perception of such phonotactic properties as syllable structure constraints. Jusczyk and colleagues provide a thorough review of these studies in the next chapter of this volume. In general, it can be concluded that infants begin at a very early age to listen selectively to the patterns of speech of their native language. In the very early months, they recognize (and prefer to listen to) the global patterns of the native language; whereas, later they begin to attend to the finer-grained structure of native-language phonetic sequences.

These rather startling findings of significant developmental changes toward language-specific patterns of speech perception in the 1st year of life altered the thinking regarding "critical" or "optimal" periods for language learning. Although few in number, studies of young children corroborated the conclusion that language-specific patterns of perception were well established long before puberty (cf. Werker and Tees 1983). The question remains whether preadolescent children are more flexible with respect to modification of phonological processing than older children and adults (see Flege this volume, and

the next section of this chapter for further discussion). In any case, the infant research refocused attention on the profound influence that early language experience has on the development of phonetic perception.

Phonetic Perception in Second Language Learners

Continued research on the perception of non-native contrasts by L2 learners corroborated earlier findings showing that adults had persistent perceptual difficulties (as well as production difficulties) with many foreign phonetic segments. For instance, in a study that employed both natural speech materials and synthetic speech continua, Mochizuki (1981) reported that Japanese learners of English residing in the United States had perceptual difficulties with the [ɹ–l] contrast, especially in syllable-initial consonant clusters. Perception of natural minimal-pair contrasts of initial [ɹ–l] was generally better than perception of a synthetic [ɹɑ–lɑ] series in which only F3 transition cues varied (see also Shimizu and Dantsuji 1983). MacKain, Best and Strange (1981) found that Japanese learners of English had difficulty perceptually differentiating synthetic [ɹɑ–lɑ] stimuli, even when both temporal and spectral cues for the contrast were present. However, Japanese learners with more English experience (including intensive conversational instruction) had significantly better performance on identification and discrimination tests than did less experienced Japanese subjects.

These studies suggest that phonetic perceptual patterns can be modified in adulthood through language immersion or intensive training on the new phonological system. However, they also suggest that, for some contrasts at least, change may be quite slow. Second language learners with years of immersion experience may still not perceive the non-native contrasts as well as native speakers. Formal language instruction may also produce only gradual perceptual changes. For example, Tees and Werker (1984) reported that after 1 year of instruction, English speakers learning Hindi in college could perceive the non-native voicing contrast but not the dental–retroflex place contrast. Students with 5 years of instruction could discriminate both contrasts.

Studies using the trading relations paradigm suggested that L2 learners may perceptually differentiate non-native contrasts on the basis of different "weightings" of acoustic parameters than those used by native listeners. For instance, Underbakke et al. (1988) demonstrated that Japanese listeners perceived the temporal distinction between AE [ɹ] and [l] even when they had difficulty differentiating the spectral differences in F3 and F2, which, for American listeners, are the primary acoustic cues. Yamada and Tohkura (1992a; 1992b) extended the study of differences between Japanese L2 learners and native English

speakers' perception of the acoustic parameters that differentiate AE [ɹ–l]. They generated stimuli in which the F2 onset/transition cue was varied independently of variations of F3 onset/transition and F1 temporal cues. Native AE speakers identified the synthetic stimuli almost exclusively on the basis of the F3 spectral and F1 temporal parameters, while F2 differences were ignored. In contrast, the Japanese listeners utilized the F2 cue to differentiate the stimuli much more often. They also identified many more of the stimuli as neither [ɹ] nor [l], but as [w], while AE listeners rarely reported hearing [w] (see Yamada, this volume, for further details of this study).

This finding was corroborated by Best and Strange (1992) who tested experienced and inexperienced Japanese learners of AE on three synthetic series: [w–ɹ], [ɹ–l], and [w–j]. (The glides [w] and [j] are phonemic in Japanese, although the Japanese [w] differs phonetically from AE [w].) Although there were no differences in performance between AE and Japanese subjects on the [w–j] series, the groups differed on both [w–ɹ] and [ɹ–l] series. Japanese listeners with less English conversational experience labeled more stimuli of the [w–ɹ] series as [w] than did the more experienced Japanese and native English speakers. Discrimination was poorer for both experienced and inexperienced Japanese listeners than for AE listeners. These results suggest that the less experienced Japanese responded more on the basis of F2 spectral differences and that the phonetic boundary on the F2 dimension was different for Japanese and English subjects. With more English L2 experience, the boundary shifted toward the English location.

The work of James Flege and his colleagues on the perception of voicing contrasts in AE syllable-final fricatives by L2 learners demonstrated non-native patterns of integration of the two temporal cues for the phonetic distinction. Flege and Hillenbrand (1986) constructed a set of synthetic "peace"–"peas" [pis–pi:z] stimuli in which vowel duration and consonant duration were varied orthogonally. (In natural speech, final [s] is distinguished by a shorter preceding vowel and longer frication noise than for final [z].) American English and French speakers, for whom the distinction is phonemic, showed a trading relation between vowel and consonant (noise) duration cues in labeling stimuli as voiced [z] or voiceless [s]. In contrast, both native Swedish and native Finnish learners of English responded only to the vowel duration cue on identification tests. This was somewhat surprising, because both Swedish and Finnish phonologies include long (geminate) and short consonants in other syllable contexts. Flege (1984) reported that Arabic learners of English with little experience also utilized only the vowel duration cue to differentiate syllable-final [s-z], whereas, Arabic learners with a great deal of English experience integrated vowel and consonant duration cues in identifying the consonants.

It appears, then, that L2 learners of English from several language backgrounds attend differentially to temporal cues for voicing contrasts in syllable-final consonants. Vowel duration differences appear to be more perceptually salient to inexperienced listeners, at least within the context of the trading relations paradigm used in these studies. (See Bohn this volume for further discussion of the relative salience of spectral vs. duration differences in vowels.)

Although most of the studies of phonetic perception by L2 learners in the 1980s concentrated on consonantal contrasts, the study by Gottfried (1984) of English speakers' perception of French vowels also demonstrated the existence of persistent perceptual difficulties with non-native vowel contrasts. American L2 learners who had studied French for an average of 7 years made significantly fewer errors on French vowels in CVC syllables than did English monolinguals with no French experience. However, L2 learners' performance was still significantly worse than native French speakers, especially on contrasts involving French front rounded vowels. In fact, their performance was no better than that of the English monolinguals on the most difficult pairs. They did show better performance on nonfront-rounded vowel contrasts on which English monolinguals had difficulties. The chapters by Bohn, Flege, and Rochet in this volume report further research on the perception of non-native vowel contrasts by L2 learners.

It can be concluded from these studies that language-specific phonetic perceptual patterns are modified by foreign language experience and, furthermore, that intensive conversational training in the L2 can facilitate perceptual learning. However, perception of difficult phonetic contrasts may improve only very gradually unless specific perceptual training is undertaken (see next section). Finally, these studies suggest that even after adult L2 learners have learned to differentiate a difficult non-native contrast, they may still perceptually integrate multiple acoustic parameters for the contrast in different ways from those used by native speakers.

Several theorists have suggested that there is a "critical" or "optimal" period for L2 learning as well as for L1 learning (cf. Krashen 1973; Scovel 1988). The perceptual and productive difficulties encountered by adult L2 learners seem to support the hypothesis that learning a new phonological system may be relatively difficult (though not impossible) after adolescence. Although anecdotal evidence about the relative ease with which younger children learn an L2 abounds, there is little empirical evidence documenting L2 learning by children of different ages, and the results are often contradictory. For instance, Shimizu and Dantsuji (1983) reported that young Japanese children perceived the AE [ɹ–l] distinction more categorically than did adult

Japanese listeners. However, Cochrane (1980) found that preadolescent Japanese children (3–13 years old) performed no better than adults on a listening test of initial [ɹ–l] minimal pairs. (They did produce the [ɹ–l] contrast better than adults.) Within the group of children, a significant (inverse) correlation between age and overall performance (perception and production) was found when sociolinguistic factors and length of exposure were held constant. However, children did not benefit from perceptual training on [ɹ–l], while adults did.

Werker and Tees (1983) reported that 4-year-old, 8-year-old, and 12-year-old English-speaking children did not differ significantly in their perception of Hindi voicing and place contrasts. All groups of children performed significantly worse than 6-month-old infants and Hindi children and adults, and no better than English-speaking adults. Thus, it would appear that even for 3- to 4-year-olds, perception of L2 phonetic contrasts may be difficult, at least initially.

Flege and Eefting (1987) examined the perception of VOT cues for stop voicing contrasts by 9-year-old Puerto Rican children who had been enrolled in an elementary school English immersion program since the age of 5 to 6 years. For 7 of the 10 children tested, the perceptual boundary between voiced and voiceless stops closely resembled that of monolingual English children; the other three had VOT boundaries intermediate between those of monolingual English and Spanish speakers. These results corroborate a study of Spanish-speaking children learning English in the United States (Williams 1979), which showed that younger L2 learners' perceptual boundaries tended to be closer to the native English boundary than were older children's after the same amount of exposure to English. Thus, it appears that immersion in an L2 environment from age 5 to 9 years may facilitate more native-like perception of voicing contrasts.

Molly Mack (1989) studied the perception of voicing and vowel contrasts by adult French-English "early" bilinguals who considered English their dominant language. All 10 subjects had acquired both languages prior to the age of 8 years from their home and school environment; four subjects acquired French first, three acquired English first, and the remaining subject acquired both simultaneously from birth. Identification and discrimination of a [d-tʰ] VOT series revealed that subjects perceived the dimension categorically with a single boundary at about +20 VOT. The location of the boundary did not differ from that of monolingual English-speaking control subjects, although identification functions for bilinguals were less steep, suggesting more inconsistency in labeling stimuli close to the boundary. Perception of an [i-I] synthetic series did show a minor difference in the location of the category boundary for bilingual versus monolingual subjects. There were no differences between bilingual children who learned French as their L1 and those who learned English as their L1. Thus, early exposure to two phonological systems appears to have had only minor effects on perception in the dominant language; that is, there was no disadvantage in early exposure to two languages.

Tees and Werker (1984) cite evidence that very early exposure to an L2 phonology may have lasting facilitative effects on the ability to learn non-native phonetic distinctions later in life. In their study of college-aged learners of Hindi, they reported that a subgroup of subjects who had been in an English-Hindi bilingual environment before the age of 2 years (with no subsequent exposure to Hindi) could perceive the difficult dental–retroflex contrast within two weeks of starting Hindi language classes. (Recall that students with no early exposure failed to perceive this contrast even after 1 full year of instruction.)

These developmental studies of L2 perception suggest that the sensitive period for phonetic perceptual learning may be considerably earlier than previously hypothesized. First language patterns of perception are well in place by 5 years of age. However, it may be that the acquisition of an L2 phonology may still be easier for preadolescent children than for adolescents and adults. Much more research is needed on the influence of early exposure to L2 phonologies on later L2 learning abilities and the effects of age of L2 exposure on phonetic perceptual patterns. Questions about the "optimum" age for L2 phonological learning are far from answered. The chapter by Yamada (this volume) addresses some of these questions for Japanese learning English as an L2.

Perceptual studies of adult L2 learners do provide encouraging evidence that, at any age, modification of phonetic perceptual patterns is possible. Second language learners with extensive immersion experience or intensive conversational training show marked improvement in the ability to differentiate perceptually even the most difficult non-native phonetic contrasts. The next section reviews recent research on the efficacy of short-term intensive training on the perception of non-native phonetic contrasts.

Perceptual Training Studies

As reviewed in an earlier section of this chapter, it can be shown that the difficulties adult L2 learners have in differentiating non-native phonetic contrasts perceptually are not because of a loss in the sensory capacity to discriminate the acoustic parameters underlying those contrasts. Early training studies also showed that adults could learn to discriminate within-category differences in phonetically relevant acoustic dimensions quite rapidly, if the training task reduced "stimulus uncertainty" and immediate feedback was provided. For example,

Carney, Widin, and Viemeister (1977) reported that AE listeners could be trained to discriminate small differences in VOT at several points along the continuum from extremely prevoiced stops to extremely postvoiced (aspirated) stops. They used a psychophysical task in which subjects judged each VOT variant against a fixed standard stimulus in a "same–different" (AX) discrimination format with feedback after each trial. Performance improved rapidly, and post-training tests showed that discrimination was a monotonic function of the difference in VOT. Carney and her colleagues also reported that two of their subjects could label VOT stimuli in terms of arbitrarily defined categories after discrimination training with the stimuli. However, these researchers did not investigate whether such training with synthetic stimuli generalized to novel synthetic stimuli or to natural speech stimuli that varied in VOT.

David Pisoni and his colleagues (Pisoni et al. 1982) also investigated the effects of training on the perception of the VOT continuum. They were explicitly interested in whether AE listeners could learn to divide the synthetic stimuli into three distinct categories—(pre)voiced, voiceless unaspirated, and voiceless aspirated—as Hindi and Thai speakers do. Training consisted of listening to repeated presentations of extreme exemplars of the three categories (-70 VOT, 0 VOT, and +70 VOT) in a fixed order, followed by 240 random presentations of the three stimuli with immediate feedback. Six of the 12 subjects reached a criterion of at least 85% correct identification after this small amount of training. Subsequent identification and discrimination tests of the entire VOT continuum revealed that these subjects divided the continuum into three distinct categories and showed two peaks of accurate discrimination for pairs that crossed the two category boundaries.

In a follow-up study, McClasky, Pisoni, and Carrell (1983) reported that 15 of 21 subjects succeeded in categorizing a [ba–p^{h}a] VOT series into three distinct classes after a brief training session. In addition, they showed transfer of training to a [da-t^{h}a] series in which VOT varied in the same way. These researchers did not assess whether trained subjects could perceptually differentiate naturally produced syllables contrasting Thai or Hindi prevoiced versus voiceless unaspirated stops.

Tees and Werker (1984) suggested that non-native voicing contrasts may be easier to train than non-native place contrasts. In their study, multiple natural tokens of Hindi CV syllables contrasting breathy voiced versus voiceless aspirated stops (distinguished by VOT) and dental versus retroflex stops (distinguished by spectral cues) were examined. Training consisted of listening to a sequence of exemplars of one category following by a sequence of exemplars from the contrasting category. Subjects determined the timing of each category change during training; a training block consisted of 50 such subject-controlled category changes.

Results of tests interspersed between training blocks indicated that 9 of 10 subjects learned to perceive the voicing contrast after only 50 training trials. In contrast, only 6 of 14 subjects reached criterion even after 300 trials (six blocks of 50 trials) on the place contrast. Furthermore, a retention test 30–40 days later revealed that although perception of the voicing contrast was maintained by trained subjects, only two of the six subjects who had successfully learned the place contrast retained accurate perception of this difficult distinction.

Strange and Dittmann (1984) also reported difficulty training Japanese L2 learners to differentiate the place contrast in AE [ɹ–l]. They employed a synthetic "rock-lock" series and the fixed-standard discrimination task that Carney et al. (1977) had successfully used to train VOT. In addition, transfer of training to different tasks, different synthetic stimuli, and to naturally produced [ɹ–l] minimal pairs was assessed in a pretest–posttest design. All eight subjects improved steadily over the course of 16 sessions in which they were trained with both an [ɹ] standard and an [l] standard in blocked trials. At the end of training, discrimination of the training stimuli was as accurate as for (untrained) native AE speakers. Seven of the eight subjects also showed significant improvement on identification and oddity discrimination tests of the "rock–lock" stimuli, and five of seven subjects showed some transfer to a synthetic "rake–lake" series (one subject was not tested). However, only two Japanese showed any improvement in perception of the minimal pair contrasts of initial [ɹ–l]. Strange and Dittmann concluded that psychophysical tasks that improved within-category discrimination were not effective in improving differentiation of non-native phonetic categories. (See Pisoni and Lively, this volume and Logan and Pruitt, this volume for further discussions of the relative efficacy of identification and discrimination training tasks.)

Donald Jamieson (Jamieson and Morosan 1986, 1989) hypothesized that successful training of non-native phonetic contrasts could be accomplished if the training task and stimuli were structured to meet three important criteria: 1) training of the contrast takes place in an appropriate acoustic (and phonetic) context, rather than in isolation, 2) an identification task (with feedback), rather than a discrimination task is used to promote phonetic categorization, and 3) sequencing of training stimuli is such that initially, attention is focused on critical acoustic parameters (by the use of extreme cases of the category), followed by the introduction of less extreme and more variable stimuli, which promotes equivalence classification of within-category exemplars. They called this procedure a *perceptual fading* technique.

Jamieson and Morosan (1986) reported success in training native French speakers to perceive the voicing contrast between the English fricatives [θ–ð] using synthetic stimuli and the perceptual fading technique. During training, synthetic exemplars were sequentially introduced starting with the most acoustically distinct exemplars and ending with a full set of synthetic tokens. Transfer of training to naturally produced fricative–vowel syllables was assessed. Results indicated that subjects rapidly learned to distinguish the fricatives and that training transferred well to novel natural speech exemplars.

In a subsequent study, Jamieson and Morosan (1989) directly compared "prototype" training (using single good instances of each of the two categories) versus perceptual fading training. Identification of the full series of synthetic stimuli was better after perceptual fading training than after prototype training. However, there were no significant differences in performance on natural speech stimuli. In yet another study (Morosan and Jamieson 1989), these researchers reported that training native French speakers on [θ–ð] in syllable–initial (CV) context using the perceptual fading technique produced improvement in perceptual differentiation that generalized to natural tokens of the fricatives in other CV syllables produced by both female and male speakers. However, there was no significant transfer of training to the contrast in medial (VCV) or syllable–final (VC) contexts. In addition, there was no indication that training on the [θ–ð] contrast improved subjects' perceptual differentiation of [ð] versus [d], even in syllable–initial contexts. Thus, it appears that subjects learned to attend to just those (abstract) acoustic parameters that differentiated the two phonetic categories in a particular syllable context. (See Rvachew and Jamieson this volume for further discussion of the limits of generalization of phonetic perception training.)

Training studies of native Japanese speakers on [ɹ–l] by Pisoni and his colleagues also suggest that perceptual learning of non-native contrasts may be specific to particular syllable contexts. Logan, Lively, and Pisoni (1991) trained Japanese on the contrast in five contexts, using an identification task (with feedback) and a large corpus of minimal-pair stimuli produced by multiple talkers. Improvement from pretest to posttest differed markedly across contexts after three weeks of training. Although all six subjects improved in their perception of [ɹ–l] in initial consonant clusters, only three subjects improved on syllable–initial [ɹ–l] minimal pairs. (Liquids in final position and in word-final clusters were identified well even before training.)

These researchers have demonstrated in several subsequent studies that identification training of Japanese subjects with a large corpus of natural speech materials leads to improvement in perception of this contrast that generalizes to novel words and to the productions of

novel speakers. However, because training materials included the contrast in all syllable contexts, it cannot be determined whether there is any transfer to novel syllable contexts. Pisoni and Lively (this volume) review much of this research and advance a theoretical framework for understanding the role of stimulus variability in perceptual learning.

In summary, recent perceptual training studies indicate that significant improvement in the perception of non-native contrasts can be achieved in a relatively short time, through intensive, focused practice with feedback. Some contrasts are easier to learn than others, but perceptual differentiation, even of the most difficult contrasts, can be enhanced. Stimulus materials and tasks that require learning to attend to phonetically relevant stimulus dimensions while ignoring (or treating as equivalent) phonetically irrelevant variations are necessary to achieve transfer beyond the specific set of training stimuli. It appears that, with the appropriate training, generalization to novel words (new phonetic contexts) and novel speakers is possible. However, transfer to different syllable positions has not yet been attained, suggesting that subjects learn to differentiate position-dependent allophones (see Flege this volume; Rochet this volume). The chapter by Rvachew and Jamieson (this volume) presents data on how procedures developed to train non-native phonetic contrasts can be applied to the treatment of phonetic perceptual problems of children with speech disorders in their native language.

Conclusions from Recent Cross-Language Research

Cross-language perceptual studies in the 1980s and early 1990s yielded a plethora of new information about the role of linguistic experience in the perception of phonetic segments. Cross-language studies investigating a wider range of contrasts and languages and employing innovative methods revealed that language-specific patterns of perception of both consonants and vowels were both robust and ubiquitous. In general, we can conclude that adults have significant difficulty perceiving most (but not all) phonetic contrasts that are not functional in their native language. That is, adults have "perceptual foreign accents" that can interfere with learning a new (L2) phonology. However, the degree of difficulty encountered by the beginning L2 learner varies significantly, depending on a multitude of factors. The psychophysical salience of the acoustic parameters differentiating phonetic contrasts, similarities and differences in the phonetic structure of L1 and L2 categories, and the phonetic and phonotactic contexts in which contrasts occur have all been shown to affect the relative difficulty of non-native phonetic categories. Thus, contrastive analyses of

L1 and L2 phoneme inventories are unsatisfactory in predicting difficulties in L2 phonological learning.

Developmental studies have clearly demonstrated that language-specific patterns of perception are established very early in the L1 acquisition process. By the end of the first year of life, selective perceptual mechanisms are attuned to only those phonetic categories, prosodic features, and syllable structures that are employed in the native language to convey meaning. Thus, native language patterns of perception are well established long before children have mastered the production of the phonetic segments and sequences of that language.

More research is needed on the question of the malleability of phonetic perceptual patterns in childhood. Studies of early bilinguals and children learning a second language in the early elementary years support the notion that the "optimal period" for perceptual aspects of L2 learning may be earlier than previously hypothesized for language learning in general. The finding that very early experience may facilitate much later L2 learning (if supported by further research) is intriguing in light of similar demonstrations of a temporal gap between the "critical period" for species-specific input and onset of production in birdsong (Marler 1975).

Research exploring the underlying bases of these language-specific patterns of perception demonstrates that psychoacoustic (sensory) capacities do not diminish as a function of maturation or experience with the native language. Adults retain the auditory perceptual abilities that are required for the detection and discrimination of the acoustic parameters that carry phonetically relevant information. Thus, we can conclude that adult L2 learners have the sensory capacity to learn new phonetic contrasts.

Experiments with adult L2 learners with differing amounts of L2 experience indicate that perceptual patterns change gradually as a function of exposure to and use of the new language. However, modification of phonetic perceptual patterns does not necessarily mirror changes in production patterns. In fact, L2 learners may actually produce non-native contrasts better than they perceive them in their own or others' speech. In this sense, L2 learning in adults follows a very different pattern from L1 acquisition. One practical implication of these findings is that foreign language teachers cannot infer perceptual mastery on the basis of assessment of pronunciation alone. It also appears to be the case that L2 learners who are able to perceive a non-native contrast may nevertheless accomplish the task in a different way than native speakers do; that is, their selective perceptual strategies may differ.

Recent perceptual training studies have reported much more success in modifying phonetic perception by adult L2 learners than earlier studies did. Significant improvement in perception even of the most difficult contrasts can be accomplished in a relatively short period of time with appropriate intervention techniques. Studies that investigate transfer of training provide important practical information about optimal training procedures. They also shed light on theoretical questions about the "psychological reality" of linguistic levels of analysis and the nature of the internal representations of phonetic categories.

In the chapters that follow in this volume, theoretical and methodological issues of current concern to cross-language speech perception researchers are discussed from a variety of points of view. In addition, important new empirical findings about the perception of speech by first- and second-language learners are reviewed. It is clear from these contributions that this is a flourishing area of research. Answers to questions about the role of linguistic experience in the perception of speech have relevance for both basic and applied scientists interested in speech and language learning and development, foreign language instruction, and speech-language disorders.

REFERENCES

Abramson, A. S., and Lisker, L. 1970. Discriminability along the voicing continuum: Cross-language tests. Proceedings of the Sixth International Congress of Phonetic Sciences. Prague: Academia.

Beddor, P. S., and Strange, W. 1982. Cross-language study of perception of the oral–nasal distinction. *Journal of the Acoustical Society of America* 71:1551–61.

Best, C. T., McRoberts, G. W., and Sithole, N. M. 1988. Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14:345–60.

Best, C. T., and Strange, W. 1992. Effects of language-specific phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*. 20:305–30.

Briere, E. J. 1966. An investigation of phonological interference. *Language*, 44:768–96.

Burnham, D. K. 1986. Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics* 7:207–40.

Burns, E. M., and Ward, W. D. 1973. Categorical perception of musical intervals. *86th Meeting of the Acoustical Society of America* JJ1, 96(A).

Burns, E. M., and Ward, W. D. 1975. Further studies in musical interval perception. *Journal of the Acoustical Society of America* 58, Suppl. No. 1: S132(A).

Carney, A. E., Widin, G. P., and Viemeister, N. F. 1977. Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America* 62:961–70.

Cochrane, R. M. 1980. The acquisition of /r/ and /l/ by Japanese children and adults learning English as a second language. *Journal of Multilingual and Multicultural Development* 1:331–60.

Cooper, F. S. 1950. Research on reading machines for the blind. In *Blindness: Modern Approaches to the Unseen Environment* ed. P. A. Zahl. Princeton, NJ: Princeton University Press.

Dalston, R. M. 1975. Acoustical characteristics of English /w, r, l/ spoken correctly by young children and adults. *Journal of the Acoustical Society of America* 57:462–69.
Eilers, R. E., Gavin, W. J., and Oller, D. K. 1981. Cross-linguistic perception in infancy: Early effects of linguistic experience. *Journal of Child Language* 9:289–302.
Eilers, R. E., Gavin, W. J., and Wilson, W. R. 1979. Linguistic experience and phonemic perception in infancy: A crosslinguistic study. *Child Development* 50:14–18.
Eimas, P. D., and Corbit, J. D. 1973. Selective adaptation of linguistic feature detectors. *Cognitive Psychology* 4:99–109.
Eimas, P .D., Siqueland, E. R., Jusczyk, P. W., and Vigorito, J. 1971. Speech perception in infants, *Science* 171:303–6.
Fant, C. G. M. 1960. *Acoustic Theory of Speech Production*. The Hague: Mouton.
Flege, J. E. 1984. The effect of linguistic experience on Arabs' perception of the English /s/ vs. /z/ contrast. *Folia Linguistica* 18:117–38.
Flege, J. E., and Eefting, W. 1987. Production and perception of English stops by native Spanish speakers. *Journal of Phonetics* 15:67–83.
Flege, J. E., and 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* 79:508–17.
Fowler, C. A. 1986. An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics* 14:3–28.
Fry, D. B., Abramson, A. S., Eimas, P. D., and Liberman, A. M. 1962. The identification and discrimination of synthetic vowels. *Language and Speech* 5:171–89.
Gibson, E. J. 1992. How to think about perceptual learning: Twenty-five years later. In *Cognition: Conceptual and Methodological Issues* eds. H. L. Pick, Jr., P. Van den Broek, and D. C. Knill. Washington, DC: American Psychological Association.
Gibson, J. J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
Goto, H. 1971. Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia* 9:317–23.
Gottfried, T. L. 1984. Effects of consonant context on the perception of French vowels. *Journal of Phonetics* 12:91–114.
Jamieson, D. G., and Morosan, D. E. 1986. Training non-native speech contrasts in adults: Acquisition of the English /ð/–/θ/ contrast by francophones. *Perception & Psychophysics* 40:205–15.
Jamieson. D. G., and Morosan, D. E. 1989. Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology* 43:88–96.
Krashen, S. 1973. Lateralization, language learning and the critical period: Some new evidence. *Language Learning*, 23:63–74.
Kuhl, P. K., and Miller, J. D. 1978. Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*. 63:905–17.
Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255:606–8.
Lane, H. 1965. Motor theory of speech perception: A critical review. *Psychological Review* 72:494–511.
Lasky, R. E., Syrdal-Lasky, A., and Klein, R. E. 1975. VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology* 20:215–25.
Lenneberg, E. 1967. *Biological Foundations of Language*. New York: Wiley.
Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. 1967. Perception of the speech code. *Psychological Review* 74:431–61.
Liberman, A. M. and Mattingly, I. G. 1985. The motor theory of speech perception revised. *Cognition* 21:1–36.
Lisker, L, 1970. On learning a new contrast. *Haskins Laboratories: Status Report on Speech Research* SR-24, 1–17.
Lisker, L., and Abramson, A. S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20:384–422.
Logan, J. S., Lively, S. E., and Pisoni, D. B. 1991. Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America* 89:874–86.
Mack, M. 1989. Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals. *Perception & Psychophysics* 46:187–200.
MacKain, K. S., Best, C. T., and Strange, W. 1981. Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics* 2:369–90.
Marler, P. 1975. On the origin of speech from animal studies. In *The Role of Speech in Language* eds. J. F. Kavanagh and J. E. Cutting. Cambridge, MA: MIT Press.
Mattingly, I. G., Liberman, A. M. Syrdal, A. K., and Halwes, T. 1971. Discrimination in speech and non-speech modes. *Cognitive Psychology* 2:131–57.
McClaskey, C. L., Pisoni, D. B., and Carroll, T. D. 1983. Transfer of training of a new linguistic contrast in voicing. *Perception &Psychophysics* 34:323–30.
Miller, J. D., Wier, C. C., Pastore, R. E., Kelly, W. J., and Dooling, R. J. 1976. Discrimination and labeling of noise-buzz sequences with varying noise lead times: An example of categorical perception. *Journal of the Acoustical Society of America* 60:410–17.
Miller, J. L., and Eimas, P. D. 1977. Studies on the perception of place and manner of articulation: A comparison of the labial–alveolar and nasal–stop distinctions. *Journal of the Acoustical Society of America* 61:835–45.
Miller, J. L., and Liberman, A. M. 1979. Some effects of later-occurring information on the perception of stop consonant and semi-vowel. *Perception & Psychophysics* 25:457–65.
Miyawaki, K., Strange, W., Verbrugge, R. R., Liberman, A. M., Jenkins, J. J., and Fujimura, O. 1975. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*. 18:331–40.
Mochizuki, M. 1981. The identification of /r/ and /l/ in natural and synthesized speech. *Journal of Phonetics*. 9:283–303.
Morosan, D. E., and Jamieson, D. G. 1989. Evaluation of a technique for training new speech contrasts: Generalization across voices, but not word position or task. *Journal of Speech and Hearing Research* 32:501–11.
Pisoni, D. B. 1973. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception &Psychophysics* 13:253–60.
Pisoni, D. B. 1977. Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America* 61:1352–61.

Pisoni, D. B., Aslin, R. N., Perey, A. J., and 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. 1991. Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America* 89:2961–77.

Polka, L. 1992. Characterizing the influence of native language experience on adult speech perception. *Perception &Psychophysics.* 52:37–52.

Polka, L., and Strange, W. 1985. Perceptual equivalence of acoustic cues that differentiate /r/ and /l/. *Journal of the Acoustical Society of America* 78:1187–97.

Polka, L., and Werker, J. F. 1994. Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance* 20:421–35.

Potter, R. K., Kopp, G. A., and Green, H. C. 1947. *Visible Speech.* New York: Van Nostrand.

Pruitt, J. S. 1992. Training native English speakers to identify Hindi retroflex–dental consonants. Unpublished Masters Thesis, University of South Florida.

Raphael, L. J. 1972. Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America* 51:1296–1303.

Repp, B. H. 1984. Categorical perception: Issues, methods, findings. In Speech and Language: Advances in Basic Research and Practice, Vol. 10, ed. N.J. Lass. New York: Academic Press.

Scovel, T. 1988. *A Time to Speak: A Psycholinguistic Inquiry into the Critical Period for Human Speech.* New York: Newbury House.

Sheldon, A., and Strange, W. 1982. The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3:243–61.

Shimizu, K., and Dantsuji, M. 1983. A study on the perception of /r/ and /l/ in natural and synthetic speech sounds. *Studia Phonologica* 17:1–14.

Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., and Öhman, S. 1969. Cross-language study of vowel perception. *Language and Speech* 12:1–23.

Strange, W. 1972. The effects of training on the perception of synthetic speech sounds: Voice onset time. Unpublished doctoral dissertation, University of Minnesota.

Strange, W. 1986. Speech input and the development of speech perception. In *Otitis Media and Child Development,* ed. J. F. Kavanagh. Parkton, MD: York Press.

Strange, W. 1992. Learning non-native phoneme contrasts: Interactions among subject, stimulus, and task variables. In *Speech Perception, Production and Linguistic Structure,* eds. Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka. Tokyo, JAPAN: OHM Publishing Co. Ltd.

Strange, W., and Dittmann, S. 1984. Effects of discrimination training on the perception of /r–l/ by Japanese adults learning English. *Perception & Psychophysics* 36:131–45.

Strange, W., and Jenkins, J. J. 1978. Role of linguistic experience in the perception of speech. In *Perception and Experience,* eds. R. D. Walk and H. L. Pick, Jr. New York: Plenum Press.

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

Streeter, L. A., and Landauer, T. K. 1976. Effects of learning English as a second language on the acquisition of a new phonemic contrast. *Journal of the Acoustical Society of America* 59:448–51.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., and Cooper, F. S. 1970. Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review* 77:234–49.

Tees, R. C., and Werker, J. F. 1984. Perceptual flexibility: Maintenance or recovery of ability to discriminate non-native speech sounds. *Canadian Journal of Psychology* 38:579–90.

Trehub, S. E. 1976. The discrimination of foreign speech contrasts by infants and adults. *Child Development* 47:466–72.

Underbakke, M., Polka, L., Gottfried, T. L., and Strange, W. 1988. Trading relations in the perception of /r/–/l/ by Japanese learners of English. *Journal of the Acoustical Society of America* 84:90–100.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., and Tees, R. C. 1981. Developmental aspects of cross-language speech perception. *Child Development* 52:349–55.

Werker, J. F., and Lalonde, C. E. 1988. Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology* 24:672–83.

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

Werker, J. F., and Polka, L. 1993. The ontogeny and developmental significance of language-specific phonetic perception. In *Developmental Neurocognition: Speech and Face Processing in the First Year of Life,* eds. B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, and J. Morton. The Netherlands: Kluwer Academic Publishers.

Werker, J. F., and Tees, R. C. 1983. Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology* 37:278–86.

Werker, J. F., and 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., and Tees, R. C. 1984b. Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America* 75:1866–78.

Williams, L. 1977a. The perception of stop consonant voicing by Spanish–English bilinguals. *Perception & Psychophysics* 21:289–97.

Williams, L. 1977b. The voicing contrast in Spanish. *Journal of Phonetics* 5:169–84.

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

Yamada, R. A., and Tohkura, Y. 1992a. Perception of American English /r/ and /l/ by native speakers of Japanese. In *Speech Perception, Production and Linguistic Structure,* eds. Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka. Tokyo, JAPAN: OHM Publishing Co. Ltd.

Yamada, R. A., and Tohkura, Y. 1992b. The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception &Psychophysics* 52:376–92.