The Production and Perception of Foreign Language Speech Sounds*

James E. Flege
Department of Biocommunication
University of Alabama at Birmingham

INTRODUCTION

This chapter has three major goals: to identify and develop theoretical issues of importance with regard to the production and perception of sounds found in a foreign language; to review systematically research dealing with these issues concluding with the year 1982 (relatively few additional studies were added after submission of this chapter in November 1983); and to identify topics of special importance for future research.

The study of phonetic aspects of foreign language learning is important for many reasons. Pronunciation is crucial to human interactions because speech carries affective and social meaning in addition to referential meaning. People seldom speak their own native language with an accent they themselves judge to be unacceptable. However, many individuals speak a foreign language with an undesired accent, or hear their native language spoken with a foreign accent. They are justifiably concerned about how a foreign accent might influence their own ability to communicate, or affect judgments made by others about them. The presence of foreign accent may have tangible negative affects on a person’s daily interactions. Moreover, foreign accent is of pedagogical importance inasmuch as it may deter someone learning a foreign language from getting the extensive practice needed for mastery of a foreign language.

An understanding of phonetic aspects of foreign language learning is also of theoretical importance to those interested in speech perception and speech motor control. Humans generally attempt to talk like those around them (or to distance themselves by pronouncing words and sounds in a way that differs systematically from those around them). To achieve this end, talkers must eventually translate patterns of sensory stimulation (auditory, visual, and tactile-kinesthetic) associated with speech gestures into stable motor plans they themselves can use to reliably reproduce those gestures. The sources of input for this gradual “transduction” process are the many thousands of words the learner encounters in the surrounding linguistic environment. The product of such experience are the many sounds, words, and sentences the learner will eventually produce in speech communication.

Previous research in a number of disciplines has documented auditory, anatomical, and physiological factors that may limit the range of sounds suitable for use in human languages. The broad outline of this research suggests that many of the fine details of speech sound production are derived from experience with specific varieties of human speech; and that the optimal perceptual processing of speech sounds results from a tuning process that derives from experience with the phonetic properties of sounds found in a specific variety of speech. Someone learning speech must develop motor control capabilities that permit a close match between output (the sounds produced) and input (the sounds perceived in the surrounding environment) to avoid sounding accented.

A long-range goal of experimental phonetics is to understand how speech processing in cortical and subcortical centers relates sensory information to stable and adaptive motor control codes used for the implementation of speech; how sensorimotor information is stored, structured, and retrieved during the production and perception of speech; and under what conditions phonetic information in central representations may be altered through experience. A better understanding of these functions may be achieved through a thorough understanding of the phonetic characteristics of foreign language learning.

This chapter focuses jointly on the production of speech sounds (i.e., context-sensitive “allophones” or phonetic “segments”) and the perception of phonetic categories. For the sake of convenience, both speech sounds and more abstract sound units (phonemes) will be represented by IPA symbols enclosed in virgules (/ /). It should be evident from the surrounding context whether a quasiphysical, phonetic represen-
The focus of this chapter should not be understood as an implicit claim that other potential sound units (such as the distinctive feature or syllable) are unimportant to speech learning. Nor is it meant to deny the potential importance of suprasegmental aspects of language (e.g., prosody and intonation) or units such as the word, to speech processing. Rather, the focus of this chapter simply reflects the reality that speech sounds are more easily observed and talked about than other sound units and have, as a result, been the object of most empirical studies of speech learning.

A major concern of this chapter is something referred to as speech learning (Locke, 1983b). Although we do not know precisely what happens when humans learn to pronounce sounds in a particular way, we assume that some kind of learning has occurred when the speaker-hearer “imposes a transformation in the processing sequence that leads to a more accurate performance on subsequent trials” (Massaro, 1975, p. 585). It seems reasonable to think that speech learning has occurred if an individual can be shown to articulate or perceive a speech sound differently after (as compared to before) massive exposure to a foreign language. Similarly, learning would seem to have occurred if the foreign language learner were found to produce or perceive native language sounds differently as the result of learning a foreign language.

Most studies to be reviewed here have employed objective measures or systematic subjective methods of observation such as phonetic transcriptions, acoustic analyses, two alternative forced-choice labelling, AX discrimination tasks. Many of the paradigms and techniques of experimental phonetics and psychology have yet to be applied to the problem of foreign language learning. However, the literature which now exists provides ample empirical evidence that speech patterns established for the production and perception of native language sounds evolve during foreign language learning, even though the magnitude of the changes observed is often insufficient to enable the learner to speak “without accent.”

Before beginning, we will first define some of the terms used most frequently in the discussion to follow. Native language is used here to refer to the first language(s) an individual learns in early childhood. The child’s principal caretaker often has a major influence on his or her developing speech, hence the alternate term “mother tongue.” The terms foreign language and target language will be used interchangeably in reference to any language(s) learned naturally or through formal instruction after establishment of the native language. The abbreviation L1 will be used to refer to the native language, and L2 (i.e., “second language”) will designate the foreign or target language. A native speaker is someone speaking her or his native language; a non-native speaker is someone speaking a foreign language.

Anyone who speaks a foreign language, whether it be the second or subsequent language(s) learned beyond early childhood, will be referred to here as an L2 learner. The term learn is used here because it emphasizes ongoing change and development. The term acquire, on the other hand, seems to connote complete mastery of some aspect(s) of L2. In the realm of speech production and perception, this is something that is seldom achieved by L2 learners.

The term bilingual will be reserved for those seemingly rare individuals who have demonstrated a native-like mastery of both L1 and L2 (although the term has been used in various other ways). An important question for future research is whether individuals who acquire simultaneously two languages in early childhood will be truly bilingual, that is, manifest a thoroughly native-like production and perception of sounds in both of their languages (see McLaughlin, 1978). Weinreich (1953, p. 9) speculated that bilinguals may develop two “coexistent” (or separate) language systems, but noted the need for experimental investigation of the possibility that “bilinguals interpret at least parts of the (two) linguistic systems as merged rather than coexistent.”

Linguists have long recognized the difficulty in distinguishing languages from dialects. Dialects of a language are speech varieties that are sufficiently distinct from one another that untrained observers may recognize their difference, yet sufficiently similar so as to be mutually comprehensible by native speakers. Languages, on the other hand, differ sufficiently to impede such mutual intelligibility. Although this chapter focuses on foreign languages, many of the phenomena examined here should, according to Weinreich (1953, p. 1), also apply to the learning of new dialect of one’s native language.

The sound system of languages may be regarded as differing at several levels of analysis. They may have a different phonetic inventory. A phonetic inventory is the ensemble of speech sounds used in speech production, the members of which are often designated using a transcriptional system such as the IPA alphabet. Languages may also differ according to the more abstract phonological inventory of phonemes used to signal differences in meaning. Two languages might, in principle, possess the same phonological inventory yet be pronounced quite differently if phonemes are implemented differently. If so, native speakers of one would be expected to produce the other with an unmistakable accent.

The sounds in the L2 phonetic inventory may be regarded as falling
into one of three broad categories: identical, similar, and new. Some L2 sounds appear to be identical to sounds found in L1. Instrumental analyses and close auditory scrutiny reveal no systematic difference between them.

Many (perhaps most) sounds encountered on the phonetic surface of L2 will differ systematically from their nearest counterpart in L1. For example, the /u/ of French and English manifest acoustic differences that are detectable auditorily; the /s/ of English and Dutch also differ acoustically even though both are transcribed with the same symbol. Although such pairs of sounds in two languages are not physically identical, but usually bear sufficient resemblance to be transcribed with the same IPA symbol. They are referred to here as similar sounds (see Flege & Hillenbrand, 1984).

The acoustic difference(s) distinguishing the similar sounds found in two languages can be regarded as deriving from language-specific realization rules that have been applied to “universal” or innately specified sound types used to implement phonetic contrasts (Chomsky & Halle, 1968). An alternative view (e.g., Flege, 1984a) is that native speakers of different languages develop different phonetic representations to motorically produce and auditorily perceive the sounds they have been exposed to during L1 acquisition (see Port & Mitleb, 1983).

Certain other L2 sounds may not bear sufficient resemblance to any L1 sound at an acoustic, auditory, or articulatory level to warrant use of the same IPA symbol. We will refer to such sounds as new, although it should be apparent that few if any L2 sounds are totally devoid of similarity to sounds in L1. For a native speaker of English, new sounds might include the /y/ of French, the pharyngeal fricatives of Semitic languages, or the clicks of Southern Bantu languages like Southern Sotho or Zulu.

Languages may also differ according limitations on the combination or sequencing of sounds. Phonological rules are thought to change the abstract structure of a phoneme before it is phonetically implemented. For example, American English but not standard German has a rule which causes /d/t/ to be realized as a flap in post-stressed, intervocalic position. Languages may also differ in the extent to which adjacent sounds are coarticulated, that is, permitted to influence the production of one another. They may also differ in the kinds of syllables used to form words. Some languages contain mostly or only CV (consonant-vowel) syllables, while other languages permit more complex syllable types (such as CVC, CCVC, or CCVCC syllables). Languages may differ in phonotactic constraints said to determine the sequence of sounds permitted to occur within each syllable type. Finally, languages manifest important prosodic and rhythmic differences which can be described as patterns of fundamental frequency, intensity, and duration affecting more than a single segment (Lehiste, 1970).

As the result of such differences, certain pairs of languages are more distant from one another than certain other pairs of languages. Researchers (e.g., Bannert, Engstrand, Eriksson, & Nordstrand, 1982) have attempted to quantify linguistic distance using formulas based on a consideration of factors such as the number of sounds which are different in L1 and L2, the magnitude of phonetic differences between L1 and L2 sounds, or the area of the differences (i.e., whether they are prosodic, pertain to the articulation of consonants, or deal with vowel production). It is generally assumed that the greater the linguistic distance between L1 and L2, the more difficult it will be for the learner to pronounce L2 like a native speaker. But, as we will see later, this is not necessarily so.

Foreign accent is a phenomenological experience of listeners that is derived from detectable acoustic (and perhaps visual) differences between native and non-native speakers in the pronunciation of sounds or other speech units. The perception of foreign accent leads to the conscious or unconscious realization by the listener that the interlocuter is not a fellow native speaker. It often (but not always) co-occurs with the overt detection of specific differences in segmental articulation that can be characterized as omissions, substitutions, or distortions. Since L2 learners with a common L1 background often produce L2 sounds so that they diverge from L1 phonetic norms in a similar way, listeners with sufficient experience can sometimes recognize the native language background of the L2 learner, as well as simply detect accent.

Sociolinguistic research demonstrates that listeners are acutely aware of differences in pronunciation, and often agree with one another concerning which variant form of a sound is the “correct” pronunciation. The term pronunciation norm refers to the collective judgment of native speakers concerning how a sound “ought” to be pronounced. Such a norm can be assessed through rating-scale judgments of acceptability and perhaps intelligibility. The phonetic norm of a language, on the other hand, is based on physical measurements of specific aspects of sound production. For example, the phonetic norm for /l/ in pre-stressed position might be partially specified in terms of voice-onset time (VOT) in ms. In English it might be 80 ms in a particular context, as compared to 15 ms in comparable speech material produced by monolingual native speakers of French.

The success with which the L2 learner pronounces L2 may be quantified according to the degree to which he or she diverges from the phonetic norms of L2. Adequacy is used here in reference to the
message perceived by the listener. For example, if an L2 learner produces the word "cat" sufficiently well for a native-speaking listener to identify it correctly in the absence of contextual semantic information, we assume that its three constituent sounds (/k/, /æ/, /t/) have been produced adequately.

Authenticity, on the other hand, refers to compliance with the phonetic norms of L2. A sound is produced adequately but not authentically when the listener correctly identifies it, but is nonetheless able to determine that it was not produced by another native speaker.

The detection of foreign accent occurs when the learner's production of an L2 sound differs sufficiently from the phonetic norms of L2. Adult native speakers, by definition, pronounce the sounds of their L1 authentically. We refer to sounds produced by native speakers which diverge significantly from the phonetic norms of L1 as distorted. Since native speakers are familiar with the types of distortions that can occur as the result of factors such as inattention or fatigue, distortions do not lead to the detection of foreign accent.

We will use the term phonetic learning to refer to changes in production and perception that occur as the result of exposure to a foreign language. Phonetic approximation refers to pronunciation changes resulting in a closer match to L2 phonetic norms, and a concomitant shift away from the phonetic norms of L1. Phonetic dissimilartion results when the learner produces L1 sounds less authentically as the result of L2 learning.

Should phonetic approximation be observed in L2 speech production, we can assume that some aspect of competence pertaining to production has also changed. Changes pertinent to perception are typically less evident than production changes to the casual observer. Such changes may result from a modification of a phonetic category prototype, the name given to the central, neurologically-encoded representation of a speech sound. Phonetic category prototypes contain detailed sensory and motor information derived from an individual's past experience producing and perceiving speech sounds. The phonetic category prototypes of monolingual native speakers of L1 and L2 are assumed to differ for similar L1 and L2 sounds that are implemented differently.

Interlanguage phonology has often been used to refer to the L2 learner's "competence" or tacit "knowledge" of the L2 sound system. As used in previous literature, these terms often refer to perceptual aspects of speech processing. It seems that most researchers would conclude that the interlanguage phonology of an L2 learner differs from that of L2 native speakers if it could be shown that the learner perceives an L2 sound (or phonetic contrast) differently than L2 native speakers. The same conclusion might not be reached, however, were it simply to be shown that the L2 learner implements an L2 sound differently (albeit recognizably), since details of phonetic realization are often regarded as falling outside the domain of "phonology" (see Flege, 1979, for a discussion).

EFFECTS OF A FOREIGN ACCENT

ACCENT DETECTION

Native speakers can often detect foreign accent in the speech of a non-native talker after hearing just a few words. Foreign accent detection is probably based on the simultaneous perception of divergences from L2 phonetic norms at the suprasegmental, segmental, and subsegmental levels (Flege, 1980; Flege, 1984a). One hears anecdotal reports concerning individuals reputed to have learned to speak an L2 in adulthood "without a trace of accent." However, the empirical evidence now available makes it appear that fewer people who learn L2 beyond early childhood (McLaughlin, 1978) actually do manage to evade recognition as non-native if their speech is scrutinized thoroughly. The evidence now available does not make it clear at which level(s) divergences are most evident to the listener, nor how the process of accent detection relates to the categorization of sounds.

Global (i.e., overall) accent has been evaluated in a number of studies. Asher and Garcia (1969) found that none of the English sentences spoken by 71 native Spanish-speaking (Cuban) children, but sentences spoken by most (23/30) native English-speaking children received the most positive (i.e., "native") rating on a 4-point rating scale of accent. The listeners in that study were native English-speaking high school students who classified incorrectly some native English-speaking talkers as non-native. In a study by Oyama (1982a), two English-speaking listeners used a 5-point rating scale to rate "phonological and prosodic" aspects of English paragraphs read by adult native speakers of Italian. Each of several groups of talkers, including those who had arrived in the United States as children and had lived there most of their lives, were judged to speak less "natively" than native English-speaking control subjects. A study by Cochrane (1977) focused on the English spoken by adult and child native speakers of Japanese. The English spoken by both adult and child subjects was judged to be more accented than that of age-matched native English controls.

Methodological studies have confirmed that global measures of degree of accent are valid and reliable. Brennan, Ryan, and Dawson (1975) had subjects estimate the degree of accent in paragraphs produced by
native and Spanish speakers of English. Degree of accent was estimated by means of magnitude estimation (i.e., assigning numbers proportional to perceived degree of accent) and by a sensory modality matching task (squeezing a hand-held device). Listeners consistently judged the speech of non-native speakers to be more accented than that of native speakers. Reliability was established by the high level of interlistener agreement, as well as the significant correlation between the magnitude estimation and sensory-modality matching tasks. Validity was demonstrated by a high correlation between the number of specific sounds that were produced incorrectly (as determined by the auditory judgments of trained listeners) and the magnitude estimation measures. A later study (Ryan, Carranza, & Moffie, 1977) demonstrated a high degree of correlation between the magnitude estimation measure of accent and rating scale judgments made by a second group of listeners.

There is increasing evidence that detection of accent may be triggered by differences between native and non-native speakers in segmental articulation. In a study by Flege (1984a), native English-speaking listeners heard short stretches of English (all or part of the syllables /tu/ and /ti/) produced by native speakers of English and French. Subjects were able to identify correctly the non-native speakers 63% to 95% of the time in forced choice and paired-comparison tasks, even though the non-native speakers' divergences from English phonetic norms were never sufficient to cue a segmental substitution.

This finding suggested that listeners do not automatically or irrevocably "filter out" the subcategorical (i.e., subphonemic) differences in segmental articulation which may distinguish similar sounds in L1 and L2 (e.g., production of /t/ with dental versus alveolar tongue contact, or with short-lag rather than long-lag VOT value). It is consistent with controlled laboratory studies (e.g., Cole, 1981) showing that listeners are able to detect mispronounced sounds that have been introduced intentionally into connected speech, despite the strong tendency in speech perception to hear phonemes that have been deleted or are masked auditorily by noise.

Another study suggested that divergences from L2 norms for stress implementation may help cue foreign accent. The ratio of stressed to unstressed vowel duration appears to be greater in English than Spanish. Hutchinson (1973) found that foreign accent judgments by native English-speaking listeners were related to the ratio of stressed to unstressed vowel duration in the English spoken by adult native speakers of Spanish. The Spanish learners who increased the duration ratio to the greatest extent were judged to pronounce English better than those who produced stressed and unstressed vowels with ratios typical of Spanish. An alternate interpretation is that changes in some other aspect(s) of

subsegmental, segmental, or suprasegmental articulation co-occurred with changes in the stressed to unstressed duration ratio, and were responsible for the improved accent ratings. However, a study employing synthetic stimuli in which only segmental duration was varied (Jonasson & McAllister, 1972) showed that listeners rejected as "foreign" those sounds produced with inappropriate durations.

It is clear that listeners can detect foreign accent easily in the speech of a non-native talker, as well as gauge its degree. However, the mechanism(s) underlying this ability is not well understood. One possibility is that the effect on accent judgments of divergences along many dimensions is additive, and that a foreign accent is detected whenever some threshold is passed. Data presented by Flege (1984a) indicated that the rate of correct detection of accent tended to be higher for relatively long intervals of speech (i.e., phrases) than short ones (single sounds), yet remained at significantly above-chance levels as the intervals were reduced from syllable-sized to segment-sized intervals, and finally to just the first 30 ms of /t/ (roughly, the burst portion of the sound /t/).

However, it is possible that accent is detected more easily in short compared to long stretches of speech not simply because there are more divergences to detect, but because attention can be more narrowly focused when the listener is exposed to short stretches of speech. Flege (1984a) observed that one French talker's production of the English phrase "two little dogs" was incorrectly accepted as native even though the first syllable in that phrase had been identified as "non-native" when presented in isolation. Thus, auditorily detectable cross-language differences in phonetic realization may not always be perceived consciously when attention is spread over many phonetic segments and suprasegmental dimensions.

Flege (1984a) hypothesized that accent detection is mediated by a comparison of sounds being processed to central phonetic category prototypes. He found that accent detection (probably unlike the recognition of accents) did not depend on prior experience. Native English-speaking listeners who were unfamiliar with French-accented English were as good at identifying French speakers of English as were listeners who were highly familiar with French-accented English. Flege interpreted this to mean that accent detection is based on perceived divergences from the phonetic information specified in phonetic category prototypes rather than on the perception of specific (non-native) properties in the speech of non-native talkers.

However, a finding by Brennan and Brennan (1981a) challenges this understanding of accent detection. Rating scale judgments and mag-
itude estimation judgments of the degree of accent in the English spoken by native Spanish speakers of English were obtained from both native Spanish- and English-speaking high school students. Each of nine Spanish talkers was assigned an “Accent Index Score” based on transcriptions of their production of 18 sounds. The Accent Index scores predicted a significant 47% of the variance in the English listeners' ratings of accent. Interestingly, they accounted for an even higher proportion of the variance (74%) in the native Spanish listeners' accent judgments.

It is unlikely that the native Spanish-speaking high school students had more highly elaborated or accurate prototypes for English sounds than the native English-speaking students. However, since they spoke both English and Spanish, they might have had greater tacit knowledge of the sound correspondences between sounds typical of standard and Spanish-accented English (see Flege & Hammond, 1982). This might have made them more sensitive to categorical and subcategorical divergences of Spanish-accented English from English phonetic norms. If so, it is possible that the Spanish-speaking listeners detected and/or identified accent at least partly by noting certain specific (non-English) properties of the speech they were processing, rather than by detecting divergences from the phonetic norms of English, as proposed earlier.

The basis for listeners' ability to gauge degree of accent is not yet understood. The simplest hypothesis is that the degree of perceived accent increases along with the number of perceived sound substitutions. Ryan, Carranza, and Moffie (1977) found that degree of accent increased along with the number of segmental substitutions in a short reading passage. Similarly, Gatbonton (1975, cited by Segalowitz & Hammond, 1982) reported a high correlation between the number and severity of substitutions and the rating scale judgments of accent provided by native English-speaking listeners.

This hypothesis is unlikely to be accurate, however, for it is probable that numerous factors influence the effect of segmental substitutions which are detected auditorily. The most obvious factor is the acoustic nature of the target and substitute sounds. Certain substitutions may be more salient than others for auditory reasons alone. Sociolinguistic factors may also influence the effect of a sound substitution. For example, Labov (1972) observed that some New York City residents stigmatize overtly the substitution of certain sounds (e.g., /d/ for /t/), as in “deese” for “these”). A speech sample containing just this substitution might be judged to be more highly accented than a sample containing a nonstigmatized substitution if stigmatization heightens listeners' awareness of a sound substitution.

The fact that /i/-for-/l/ substitutions were produced twice as frequently as /u/-for-/u/ substitutions suggested that the subjects were more aware of the former than the latter. This interpretation is consistent with results reported by Brennan & Brennan (1981b), who found that the frequency of /i/-for-/l/, but not /u/-for-/u/, substitutions was directly related to how negatively native English speakers rated native Spanish speakers of English on status dimensions such as “wealth” and “level of education.”

It might be argued that subjects produced certain substitutions as the result of reactivating some of the phonological processes said to “simplify” segmental articulation in child speech (see Flege & Davidian, 1983, for a discussion). For example, the subjects may have substituted /til/ for /ll/ not because they knew tacitly that such a substitution occurs in Spanish-accented English, but because /ll/ is easier to produce than /til/ (see Michaels, 1974). This seems unlikely, however, since the subjects produced few “simplifying” substitutions that are not actually heard in Spanish-accented English.

Another possibility is that the frequency differences reflected the actual occurrence of sound substitutions in Spanish-accented English. The substitution of /i/-for-/l/ may have occurred more frequently than /u/-for-/u/ substitution because the subjects had heard the former substitution pattern more frequently, and therefore remembered it better. If so, the /i/-for-/l/ substitution might not be a stronger indicator of Spanish accent.

It would be useful to test directly the hypothesis that certain substitutions influence degree of accent to a greater extent than others. To do so, trained native English-speaking talkers unaware of the purpose
of the experiment could be asked to read similar passages with an equal number of /i/-for-/I/ and /u/-for-/u/ substitutions. The task of subjects would be threefold: to evaluate degree of accent in the passages, make subjective evaluations of the talkers, and note specific aspects of nonstandard pronunciation. The hypothesis that certain substitutions are stigmatized would be supported if the subjects report hearing /I/-for-/I/ and /u/-for-/u/ substitutions with equal frequency, but give lower ratings and downgrade talkers to a greater extent as the result of the former substitution pattern.

Another possibility is that stigmatized substitutions are less easily detectable than nonstigmatized substitutions because they are more predictable. Unfortunately, at present we have no certain knowledge concerning the effect of familiarity on subjects' evaluation of sound substitutions. Adult native speakers may actually hear “rabbit” when children say “wabbit” because children’s substitution of /w/ for /r/ is so common. The ability to learn patterns of correspondence between sounds in two varieties of speech (Lovins, 1976) seems to be the basis for why the English spoken by non-native speakers and unfamiliar children becomes more intelligible after a brief period of familiarization. It is possible that the speech errors of children do not lead to the perception of accent because they are often predictable.

Segmental substitutions in the speech of adult (non-native) learners may be relatively less predictable because mature native speakers usually have less contact with non-native speakers than native-speaking children, and because of the wide range of possible foreign accents. This leads to the hypotheses that listeners are less aware of predictable than unpredictable sound substitutions, and that unpredictable substitutions lead to a higher degree of perceived accent (and greater social downgrading) than predictable ones. Still another hypothesis is that an L2 learner’s accent will appear to improve as native speaking listeners become familiar with it. This hypothesis received preliminary support from a study by Gass and Varonis (1984).

An important question about sound substitutions is whether phonemic substitutions affect accent judgments to a greater extent than sub-phonemic ones. Phonemic substitutions are segmental substitutions that result in a perceived change in speech sound category, whereas sub-phonemic substitutions do not. For example, the implementation of /u/ as an /y/-quality vowel in an English word would not lead native English speakers to perceive the substitution of one vowel category for another, whereas the implementation of /u/ as an /o/-quality vowel would have that effect (see Gottfried, 1984; Flege 1984a).

Evidence pertinent to this question was provided by Johansson (1978). Johansson, a native speaker of Swedish, produced English sentences with some of the sound substitutions typical of Swedish-accented English. The phonemic substitutions were often, although not invariably, judged to be more accented than the subphonemic substitutions. For example, a sentence containing the substitution of a uvular for tongue-tip /r/ was judged to be more accented than one containing an /s/ for /θ/ substitution (as in “somesing”). The various sub-phonemic substitutions received a wide range of accent ratings. The reason for this, Johansson speculated, was that some substitutions typical of Swedish-accented English also occur in certain English dialects and speech styles, thereby making them sound less “foreign” to English-speaking listeners because of their familiarity.

The phonemic errors also received a wide range of accent ratings. For example, substitution of /ε/ for /æ/ in “Land’s End” (yielding “lend’s end”) was judged to be more accented than the reverse substitution pattern (yielding “Land’s and”). The name “Ted” received relatively high ratings when produced with an /æ/-quality vowel probably because, in this instance, an /æ/ for /ε/ substitution results in another equally acceptable English name (i.e., “Ted”).

This last finding leads to the hypothesis that the effect of segmental substitutions varies as a function of semantic context. This hypothesis is in keeping with the general facilitation for feature detection by the presence of context which characterizes subjects’ conscious reports of sensory experiences (e.g., the detection of phonemes in words versus nonwords; see Rubin, Turvey, & Van Gelder, 1976). If so, we might expect that a change in segmental articulation would be detected more rapidly, and lead to a higher degree of perceived accent, when it occurs in a word that is predictable from context than in the same word presented in isolation.

Finally, it is likely that the listener’s attitude towards his/her interlocuter will influence importantly the effect of any categorical or subcategorical divergence. Giles, Taylor, & Bourhis (1973) tested the hypothesis that listeners’ subjective attitudes are influenced importantly by their perception of how hard the interlocuter is trying to accommodate to them linguistically. A French Canadian recorded a text in several different guises: French, English mixed with French, comprehensible but disfluent English, and fluent but clearly French-accented English. English Canadian subjects who could speak French were randomly assigned to one of the four guises. Their task was to return a message in response to what they had heard on the tape. The native English subjects showed the greatest evidence of accommodation (i.e., responding entirely or partly in French) to the disfluent-but-intelligible guise. This suggested they evaluated that guise more favorably than the other guises because they perceived the talker as trying hard to
accommodate to them. If so, it is possible that the negative effect of (remaining) divergences will actually increase as an L2 learner gains greater proficiency in L2 because native-speaking listener's perceive him or her to be making less effort to accommodate.

**DIMINISHED ACCEPTABILITY**

A number of studies have attempted to evaluate how “good” or how “acceptable” L2 pronunciation is. Sounds judged to be accented are generally judged to be less acceptable than unaccented sounds, but it has never been demonstrated that “acceptability” and “accentedness” represent a single dimension. It is at least logically possible that a sound might be judged acceptable and accented. We will therefore consider studies which have evaluated acceptability and those examining degree of accent separately.

One difficulty in assessing degree of accent is that L2 learners often make both grammatical and pronunciation errors, and errors in one domain might influence judgments concerning adequacy in the other. Two studies attempted to disentangle the effects of these possibly separable dimensions on accent judgments. Varonis and Gass (1982) had an accented non-native speaker of English produce grammatical and ungrammatical versions of the same sentences. Not surprisingly, the grammatical sentences were judged “good” more often (about 70% of the time) than the ungrammatical sentences (40%). The authors hypothesized that grammaticality affects the evaluation of an L2 learner’s pronunciation only when it falls in the mid-range of acceptability. This hypothesis tended to be confirmed by the results of a rating scale evaluation of accent in the same speech samples. Ensz (1977), using the matched guise technique, presented samples of French produced with a varying number of pronunciation, grammar, and vocabulary errors by native English speakers. Responses to a semantic differential scale suggested that all three components were important, but none to the exclusion of the others.

Elsendoorn (1984) reported rating scale data pertinent to the acceptability of isolated CVC English words produced by a native speaker of English and Dutch. Native English-speaking listeners accorded higher ratings to the English than Dutch native speaker, and gave the Dutch native speaker higher ratings for words ending in /p,t/ (sounds found at the end of Dutch words) than /b,d/ (sounds which do not terminate Dutch words). Elsendoorn used a digital waveform editor to replace sounds produced in CVC words by the native English speaker with sounds produced by the Dutch speaker. Acceptability judgments were higher when the initial consonant was Dutch-produced than just the final consonant; and considerably higher when just a consonant (in initial or final position) than just the vowel was Dutch-produced. This suggested that divergences from L2 phonetic norms for consonants affects acceptability more than divergences from the phonetic norms for vowels, even though both kinds of divergences may be equally detectable (Flege, 1984a).

There is conflicting evidence concerning the effect of segmental timing on acceptability judgments. Elsendoorn (1983b, 1984) changed the duration of vowels produced by Dutch native speakers in English CVC words to match the average duration of vowels produced by native English speakers in the same words. This temporal manipulation had little apparent effect on acceptability judgments, leading Elsendoorn to conclude that departures from spectral phonetic norms may influence acceptability judgments to a greater extent than divergences from the temporal phonetic norms of L2.

An argument against this interpretation is that some spectral divergences from L2 phonetic norms seem to have resulted in a segmental substitution (i.e., /e/ for /æ/) which might, in itself, have led to lower acceptability ratings. However, it should be noted that negative judgments did not depend exclusively on the perception of segmental substitutions. When the duration of the relatively short Dutch-produced vowels in words ending in /b/ and /d/ were lengthened (so as to equal the average duration of vowels produced in this context by native English speakers), the frequency of perceived /p/-for-/b/ (and /t/-for-/d/) substitutions decreased about 20% without a concomitant improvement in acceptability.

A study by Jonasson and McAllister (1972) suggested that the temporal specification of vowels and consonants does have an important effect on acceptability judgments. The durations of vowels and consonants in a naturally produced Swedish word were manipulated by computer editing techniques. Twenty Swedish-speaking subjects made categorical judgments and acceptability judgments using a 3-point rating scale. The listeners gave the highest ratings to stimuli whose temporal specification most nearly conformed to the segment durations measured in spoken Swedish. Using these judgments as a metric, predictions were made concerning the acceptability of Swedish words produced by a native and non-native (American) speaker. Over 70% of the Swede's words would have been given the highest rating, compared to less than 30% of the American's.

The reason for the seeming discrepancy between Jonasson and McAllister (1972) and Elsendoorn (1983b, 1984) reports is unclear. One possibility is that the central phonetic representations for vowels specify a narrower range of acceptable values in Swedish than English. In
English, vowel duration has an important influence on categorical identity only for vowels of spectrally ambiguous quality. In Swedish, on the other hand, vowels of the same quality may be identified differently if they differ in duration.

Two other studies examined the phonetic implementation of stress. Willems (1982) observed that the fundamental frequency differences between accented and unaccented syllables is greater in English than Dutch. Dutch speakers tended to produce accented syllables in English words with an Fo rise or fall that was far smaller in magnitude (1–7 semitones) than the Fo change seen for native speakers of English (4–12 semitones). Acceptability judgements by native English-speaking listeners increased as the pitch contour in synthetic sentences increasingly resembled the values actually produced by native English speakers. Experiments with resynthesized natural speech suggested that many of the syllables typically produced by Dutch speakers of English would be judged unacceptable if judgments depended just on Fo.

**DIMINISHED INTELLIGIBILITY**

One consequence of nonauthentic pronunciation is diminished intelligibility, which is usually quantified as the percent correct recognition of words produced in isolation or in fixed carrier sentences. Several studies have shown that the speech of L2 learners is less intelligible than that of native speakers, at least under conditions of distortion. Lane (1963) presented lists of words produced by a native speaker of English, Serbo-Croatian, “Indian” (sic) and Japanese to native English-speaking listeners. The speech of the non-native speakers was 36% less intelligible, on the average, than that of the English native speaker. Decreasing signal-to-noise ratios and frequency bandwidths resulted in similar decrements in intelligibility for the speech produced by both the native and non-native speakers. Nabèléck and Donahue (1984) found that the English spoken by 20 non-native speakers was as intelligible as that of 20 native speakers of English under normal listening conditions, but significantly less intelligible in reverberant (T=0.4, 0.8, 1.2 sec) conditions (see also Bergman, 1980). S. Johansson (1978) presented English sentences produced by native and non-native speakers of Swedish to English-speaking listeners. Comprehension (as measured by the subjects’ ability to complete sentences) was much higher for the speech samples produced by the native than non-native speakers. Variability in L2 speech may lead to decreased intelligibility. Gupta and Mermelstein (1982) programmed a speech recognition device to recognize a set of words produced by native and non-native speakers. The device correctly recognized more of the words produced by native than non-native speakers. This seems to have occurred because the native speakers resembled one another to a greater extent than the non-native (i.e., French Canadian) speakers, who probably approximated the phonetic norms of English to varying extents. As a result, it was easier for the device to construct the templates needed for word recognition for the native than non-native speakers.

It is likely that many factors in addition to divergences from L2 norms for segmental and suprasegmental articulation and increased variability will influence the intelligibility of L2 speech. Monsen (1983) found that the intelligibility of English spoken by deaf people varied considerably (from 57% to 96% correct) as a function of listeners’ familiarity with “deaf accented” English, the presence/absence of semantic context, the presence/absence of visual cues, the complexity of syllable structure, and the number of prior repetitions. Gass and Varonis (1984) found that the English spoken by L2 learners was more intelligible to listeners when preceded by meaningful context than not, and that the intelligibility of L2 speech increased as listeners grew familiar with it. This suggests that, as for deaf speech, no single test will provide an adequate assessment of L2 learners’ speech intelligibility.

**NEGATIVE EVALUATION**

Perhaps the most important aspect of foreign accent is its effect on how the L2 learner is perceived as a person. Most existing language attitude research has examined listeners’ reactions to the speech of talkers differing in ethnic, regional, or social background. The “matched guise” technique developed by Lambert and his colleagues (e.g., Lambert, Hodgson, Gardner, & Fillenbaum, 1960) requires the listener to evaluate a talker subjectively based on his or her pronunciation. The listener hears a talker producing the same standard passage in various different “guises” (that is, talking with a different social dialect or accent), but is unaware that the passages have been produced by a single talker. A few studies have examined reactions to speech produced by different groups of talkers.

The general finding is that a talker who speaks like members of the highest socioeconomic class (the so-called “prestige dialect”) will be evaluated more positively on dimensions relating to social status (e.g., estimated professional success, education, intelligence, wealth) when he or she speaks like members of the prestige dialect than when variants of a non-prestige dialect are used. That is, talkers who appear to represent a non-prestige dialect tend to be downgraded compared to talkers representing the prestige dialect, even by listeners who are themselves native speakers of the non-prestige dialect. On dimensions
relating to more personal qualities such as trustworthiness, friendliness, and kindness, on the other hand, talkers who appear to represent a non-prestige dialect may be downgraded relatively less by fellow native speakers of non-prestige dialects. They may even receive more positive evaluations than talkers who appear to represent the prestige dialect.

A number of studies (e.g., Anisfeld, Bogo, & Lambert, 1962) indicated that individuals who speak with a foreign accent may also be downgraded severely, although native speakers seem to be more tolerant of non-native children than adults (Lepicq, 1980, cited by Ryan, 1983). According to Ryan (1983, p. 8), talkers who speak with a foreign accent are regarded as "uneducated, unintelligent, and relatively poor, especially if the (accent) elicits an association with a lower class minority group." Supporting this is a finding reported by Kalin and Rayko (1978), who had native English speakers rate native and non-native speaker's suitability for jobs of varying status. The evaluations were based on 30-sec excerpts of "informal, colloquial English that was grammatically correct" (p. 1205). To control for content, a native speaker reproduced the speech sample that had been produced by a non-native speaker, and vice-versa. Five non-native speakers (Italian, Greek, Portuguese, West African, Yugoslav) were generally judged to be less suitable for high status jobs than native speakers, but received higher ratings than native speakers for low-status jobs. Perhaps predictably, subjects shown by personality tests to have authoritarian and ethnocentric personality characteristics tended to discount the suitability of the non-native speakers for high-status jobs more than those who did not have these traits. In another study (Hinofotis & Bailey, 1981) listeners evaluated the teaching performance of non-native teaching assistants in American universities. Both undergraduate students and college personnel judged pronunciation to be more important than vocabulary, grammar, speaking rate, clarity, and even the logical development of ideas!

According to some researchers (e.g., Lambert, 1967) the negative evaluation accorded speakers of non-prestige dialects depends on a two-stage process. First, the listener assigns the talker to a specific social group, then associates learned group stereotypes to the talker. However, there are several reasons to think this two-step process may not apply to the evaluation of foreign-accented speech. Ryan and Carranza (1975) found that listeners' evaluations of non-native speakers changed as a function of the context in which the speech was thought to have been produced (i.e., school versus home). Ryan and Sebastian (1980) found that evaluations were more positive when listeners were led to believe that non-native talkers were middle-class rather than members of a lower socioeconomic class.

The evidence which now exists suggests that subjective evaluations based on foreign accent tend not to fall into binary categories (detected or not detected). If negative evaluations derive from the evocation of learned stereotypes, they should depend only on a correct recognition of group membership, and be unaffected by information irrelevant to group membership. And, if subjective evaluations depend on learned stereotypes they should not vary continuously. What has been found, instead, is that evaluations do seem to vary continuously according to degree of foreign accent which, in turn, probably depends on the extent to which non-native speakers approximate L2 phonetic norms. Ryan et al. (1977) examined listeners' evaluations of the status (e.g., eventual occupation) and personal worth (e.g., likelihood of friendship) of non-native speakers. Their judgments became less positive as degree of accent increased. Brennan and Brennan (1981a) also found that the extent to which talkers were downgraded on status dimensions (e.g., educated versus uneducated; wealthy versus poor; successful versus unsuccessful; intelligent versus unintelligent) depended on degree of accent.

Although several studies have indicated that foreign accent leads to negative evaluation, other studies have suggested that foreign accent has no absolute or inherent effect, but depends on the listener's overall perception of the talker. As mentioned earlier, evaluations can be affected by the context in which foreign-accent speech is produced (Ryan & Carranza, 1975) as well as the perceived social class of the non-native talker (Ryan & Sebastian, 1980).

In another study (Nisbett & Wilson, 1977), students were shown a videotape of a professor who spoke with a foreign accent. One group saw him behaving in a warm and supportive manner, while another group saw him behaving in a cold and irritable manner. The students, of course, evaluated the professor more positively in his "warm" than "cold" guise. Interestingly, students who evaluated the professor positively cited his accent as a positive attribute while those who reacted negatively to the professor cited his accent as a negative attribute. This is consistent with the fact that TV advertisements sometime employ actors who speak with (or feign) a foreign accent. This probably would not be done unless a foreign accent is perceived in at least some contexts to be desirable.

In summary, we have seen that in many circumstances foreign accent may have a negative effect on how L2 learners will be evaluated. As we will see in the next section, few individuals ever learn to speak L2 without an accent. Therefore it is important to understand why listeners tend to react negatively to foreign-accented speech. Several hypotheses have been offered in the literature.

First, foreign accented speech may be downgraded because it is
difficult to understand (Sebastian, Ryan, Keogh, & Schmidt, 1980; cited Ryan, 1983). Difficulty in processing the speech of a non-native might cause discomfort ("cognitive stress") to the listener.

Second, some of the substitution patterns seen in L2 speech (e.g., /t/ and /d/ for the initial fricative of "thank" and "the") may resemble those found in nonstandard and lower-class dialects of L2. They may, therefore, tend to evoke the same negative evaluations they would evoke had they been produced by a native speaker.

Third, listeners might develop stereotypes concerning foreign-accented talkers in social contexts where native speakers of different languages live side by side (e.g., Quebec). It is possible that in such a social context a variety of foreign-accented speech might function as a social dialect. Gatbonton (1975, cited by Segalowitz & Gatbonton, 1977) found that native L2-speaking listeners made inferences about the attitudes of L2 learners based on their pronunciation of L2. Native English speakers tended to believe that French Canadians who pronounced English with a relatively high degree of authenticity thought favorably of native speakers of English. They also thought that French Canadians who spoke English with a strong accent were less kindly disposed toward native English-speaking Canadians. Another finding was that French Canadians tended to downgrade fellow French Canadians who spoke English authentically as less suitable for leadership roles than French Canadians who spoke English with an strong French accent (but cf. Hill, 1970). Similar results are reported for Canadian children by Genesee and Holobow (1978).

The hypothesis that listeners develop stereotypes for individuals who speak with a foreign accent should not be accepted uncritically. For example, the possibility exists that listeners develop a stereotypic reaction toward any nonstandard form of speech which is not recognizable as a dialect of their own native language. I know of no empirical study which has assessed the ability of listeners to recognize specific varieties of foreign-accented speech (e.g., Italian-accented English). This ability may be difficult to achieve. Identification of ethnic, social, and regional dialects is possible because they are relatively stable forms of speech with which listeners have extensive experience. However, listeners ordinarily experience many different types of foreign accent, and the foreign accent of any individual L2 learner is apt to diminish as he or she gradually approximates the phonetic norms of L2.

FACTORS CAUSING ACCENT

A child is said to "master" the sounds of L1 over the course of several years. Do children eventually learn to produce the sounds of L1 authentically because they are in a special state of readiness for speech acquisition? Because of their extensive experience producing and perceiving the sounds of L1? Or simply because they gradually develop the motor control necessary for the precise and reliable articulation of speech sounds? An obstacle to answering such questions is the inevitable confounding of experience and speech motor development.

Similar questions are asked in regard to L2 learning. Since most L2 learners can produce L1 sounds authentically by the time L2 learning begins, motor control is not as important an issue for adult L2 learning as it is for the early stages of L1 acquisition (Smith, 1978). However, researchers ask "Does L2 pronunciation depend more importantly on the age at which L2 learning first begins, or on the amount of experience in L2?" Although age of first exposure to L2 and the amount of L2 experience have been confounded in some studies, researchers are attempting to untangle these two variables. Age of first exposure to L2 appears to be important to L2 speech learning, although this factor may be confounded with differences in the amount and perhaps quality of L2 input.

AGE AND AMOUNT OF L2 EXPERIENCE

A number of studies using rating-scale judgments of accent by both trained and untrained listeners have indicated that the authenticity of a learner's pronunciation varies as a function of the age at which he or she was first exposed to L2. Asher and Garcia (1969) presented English sentences spoken by children whose native language was English or Spanish to native English-speaking listeners. Most (23 of 30) native English-speaking children were correctly identified as "native"; none were identified as "definitely accented." Not one of 71 native Spanish-speaking children was identified as "native," although children who arrived earliest in the United States (1-6 years of age) received somewhat higher ratings than children who arrived later (7-12 and 13-19 years of age). Similarly, Oyama (1982a) examined the English spoken by 60 Italian-born men. There was a strong negative correlation between age of first exposure and degree of accent. Subjects who arrived in the United States between 6-10 years of age were judged to have better overall accents than subjects who arrived between the ages of 11-15 years. In turn, these subjects were judged to pronounce English better than subjects who arrived between the ages of 16-20 years.

Studies examining the production of specific sounds rather than assessing global accent have also provided evidence that early exposure to L2 promotes authenticity of pronunciation. Williams (1979, 1980) found that, with length of residence in the United States held constant,
native Spanish-speaking children aged 8 to 10 years approximated more closely the phonetic norms of English for stop consonant production (as determined by VOT measurements) than children 14 to 16 years of age. Snow and Hoefnagel-Höhle (1977) examined the pronunciation of 47 adults and children learning Dutch naturalistically over a period of one year. There were few differences between subjects of different ages in the imitation of Dutch vowels and consonants in the first data sample. There was, however, a strong positive correlation between age and authenticity of pronunciation in a spontaneous picture-naming task. This runs counter to the general belief that children pronounce L2 sounds better than adolescents or adults. The gap between the older and younger learners grew smaller in two subsequent data samples, in which the expected negative correlation between age and authenticity finally did emerge for a number of Dutch sounds. The authors suggested that children's L2 pronunciation success may ultimately derive from the fact that they go on learning longer than adults, rather than from a greater ability to learn new forms of pronunciation.

The age of first exposure to L2 and the amount of L2 experience have often been confounded in past studies. However, it seems that age of first exposure to L2 represents a factor that may influence L2 pronunciation success independently of amount of experience. Fathman (1975) used a rating scale to evaluate the English spoken by two hundred non-native children learning English in American public schools. Children 6 to 10 years of age received significantly higher ratings than children aged 11 to 15 years, even when length of residence in the United States was held constant at one, two, or three years. Oyama (1982a) found that the effect of age of first exposure remained significant when the effect of length of residence in the United States was statistically removed in partial correlation analyses.

Two studies which examined more abstract components of pronunciation suggest (contrary to evidence concerning the pronunciation of specific sounds) that older learners may be superior to younger learners in making generalizations about how sounds are pronounced. Ervin-Tripp (1977) noted that older children learn phonological rules faster than younger children. Cochrane and Sachs (1979) found that, although children outperformed adults in imitating Spanish sounds, there was no difference between children and adults in abstracting the rule for stress placement in Spanish words on the basis of a limited exposure to Spanish words.

Taken together, these results seem to support the belief that children often pronounce an L2 better than adults. They also support the conclusion that the age at which L2 learning commences affects L2 pronunciation, at least for children and adolescents learning an L2 naturalistically. Age differences among adult L2 learners seem to have little or no effect.

It is less certain whether amount of L2 experience can be used to predict a learner's success in pronouncing L2. Asher and Garcia's (1969) data indicated that Cuban children who had lived 5 to 8 years in the United States pronounced English more authentically than children who had lived only 1 to 4 years in an English-speaking environment. Tahia, Wood, & Lowenthal (1981) used a 3-point rating scale to evaluate the pronunciation of an English paragraph read by 115 non-native speakers who had lived at least two years in the United Kingdom. The age of first exposure to L2 accounted for a significant 43 % of the variance in a multiple correlation analysis, but length of residence in the United Kingdom did not predict L2 pronunciation success. The nonsignificance of amount of experience may have been due to the fact that subjects who had lived less than two years in the United Kingdom were excluded from the study. It is during this period that authenticity of L2 pronunciation is likely to improve most (see, e.g., Asher & Garcia, 1969).

Cochrane (1977) examined the ability of Japanese children and adults to produce and perceive the contrast between English /r/ and /l/. Amount of experience (based on the number of years of residence in the United States and subjects' self-report concerning how many hours per day they spoke English) was not found to predict success in producing and perceiving /r/ and /l/. However, once the effect of differences in motivation and amount of experience was partialled out using statistical techniques, the expected negative correlation between age of arrival and L2 pronunciation emerged for the 54 child subjects, who were 1 to 12 years of age when first exposed to English.

Flege and Davidian (1985) found that length of time in the United States did not affect significantly learners' production of stop consonants in word-final position. A total of 36 L2 learners (native speakers of Spanish, Chinese, and Polish) who arrived in the United States during late adolescence or adulthood were perceived to devoice English /b,d,g/ about 40% of the time. Only a nonsignificant 14% of the variance in the frequency of devoicing scores was accounted for by the talkers' length of residence in the United States and self-reported frequency of use of English.

Many instrumental phonetic studies have shown that the amount of L2 experience has a subtle, yet quantifiable influence on certain aspects of segmental articulation. Most of these studies have compared groups of adult L2 learners with similar educational and social backgrounds who differ according to length of time spent in an L2-speaking environment. For example, several studies to be reviewed later have shown that experienced, but not relatively inexperienced, L2 learners
resemble native English speakers in producing /p,t,k/ with a longer period of supraglottal stricture than /b,d,g/ in word-final position. Flege and Hillenbrand (1984) examined the spectral quality of the French vowel /u/ produced by undergraduate students who had spent the previous academic year in France, and professors of French who had also lived at least one year in France. Acoustic and perceptual measures revealed that the experienced native English speakers approximated the phonetic norms of French for /u/ more closely than those with relatively less L2 experience.

An effect of experience has not been evident in instrumental studies investigating other acoustic dimensions in L2 speech, however. These studies have shown that neither inexperienced nor relatively experienced L2 learners make vowels as much longer before /b,d,g/ than /p,t,k/ as native speakers of English. Moreover, amount of experience seemed to have little effect on the extent to which adults approximate the L2 phonetic norm for VOT in pre-vocalic stops (Suomi, 1980; Port & Mitleb, 1983; Flege & Hillenbrand, 1984). For example, Flege and Port (1981) found that a group of Saudi Arabians who had lived for an average of nine months in the United States produced English stops with the same VOT values as Saudi Arabians who had lived there an average of 39 months. Both groups produced /p,t,k/ with VOT values that were significantly shorter than those produced by native English speakers. However, amount of experience may affect children's production of L2 stops. Williams (1979, 1980) found that both 8 to 10 and 14 to 16 year-old Puerto Rican children who had lived 3-3.5 years in the United States produced /p/ with longer VOT values (i.e., stops that were more English-like by about 10 ms) than children who had lived less than 6 months in the United States. It was evident, however that with length of residence in the United States held constant, the younger children produced /p/ with VOT values that were 5-25 ms longer (and therefore more English-like) than the older children.

Taken together, these studies suggest that amount of experience has relatively little effect on L2 pronunciation. Little or no effect of amount of experience is observed when L2 pronunciation is assessed at a global or segmental level. However, instrumental studies have shown that L2 phonetic norms for certain specific aspects of segmental articulation may be more closely approximated by subjects with greater L2 experience than those with less L2 experience.

THE CRITICAL PERIOD HYPOTHESIS

How might we explain the seeming advantage of early as opposed to later L2 learning? Some investigators have concluded that children are inherently better than adults in learning new forms of pronunciation because they have not passed a “critical” period for language learning. Penfield and Roberts (1966) asserted that cortical centers important for language acquisition lose “plasticity” by about 12 years of age. Lenneberg (1967, p. 176, 377) concluded that “language readiness . . . begins around two and declines with cerebral maturation in the early teens.” He speculated that adults inevitably speak foreign languages with an accent if L2 learning begins after childhood because the ability to learn new forms of pronunciation is inhibited once cerebral processes become “firmly structured” as the result of lateralization. Lamendella (1977, p. 165) argued that the “immature neurolinguistic systems” of children give them “intrinsic greater potential” for L2 learning than adults. Scovel (1969, p. 245) claimed that:

It is the nature of the human brain, not its nurture, that crucially affects (L2 pronunciation). The onset of cerebral dominance, which seems to occur around the age of twelve, inhibits the ability of a person to master the sound patterns in a second language without an impinging foreign accent.

Walsh and Diller (1981, p. 12) provided specific new arguments concerning a biological basis for “difficulty in conquering foreign accents after childhood.” They noted (1981, p. 16) that although L2 learners may thoroughly acquire aspects of a foreign language other than pronunciation (e.g., vocabulary), complete success in pronouncing L2 is impossible because pronunciation is a “lower order” linguistic function which is “genetically specified and consolidated in early development.” They speculated that pronunciation patterns are based on early maturing “neural circuits,” whereas the development of “higher order” linguistic components such as the lexicon depends on stellate cells, which are largely undifferentiated at birth and continue to develop long after birth. According to Walsh and Diller, stellate cells play an important role in the establishment of new neuronal synaptic arrangements, are associated with the “plasticity” discussed earlier by Lenneberg (1967) and Penfield (1965, Penfield & Roberts, 1966), and are important to various aspects of learning and memory. According to Walsh and Diller (1981, p. 8), the auditory processing of speech sounds may depend on the existence of “innate neural detectors for phonological distinctive features” or neuronal circuits that are established early and do not evolve with experience.

Walsh and Diller’s hypothesis concerning speech perception is clearly inconsistent with evidence to be reviewed later. However, the possibility for a neurally based loss of “plasticity” finds prima facie support in
the hormonally-induced loss of ability for further vocal learning observed in certain song birds known as "crystallization" (Marler & Mundinger, 1971; Studdert-Kennedy, 1981), as well as in the oft-cited superiority of children compared to adults in L2 pronunciation.

The "critical period" hypothesis makes two important predictions regarding the pronunciation of foreign languages. First, to be entirely effective, language acquisition must occur prior to the establishment of hemispheric specialization for language functions, something which is thought by some to occur at about the age of 12 years. Second, language learning that occurs after the critical period has been passed will proceed more slowly, and ultimately be less successful, than learning observed in normal L1 acquisition (which is assumed to occur before the critical period has been passed).

There are numerous reasons for questioning the hypothesis regarding a critical period for speech learning (at least as it was just formulated). To begin, the concept was developed in ethological studies examining behavior in non-human rather than human species. Four characteristics appear to define behavioral pattern whose acquisition is limited by a critical period:

1. It tends to appear under well-defined developmental conditions.
2. It cannot be forgotten or revised once it has been established.
3. It involves the recognition of specie rather than individual characteristics.
4. It may be learned long before it is manifested.

A good example of a "critical period" for the learning of environmentally important stimuli is the imprinting seen in certain avian species. Of the four characteristics just mentioned, only the first seems to apply directly to human speech learning.

Oyama (1979) argued that the critical period concept may be useful for delineating developmental phases and specific mechanisms used to regulate sequences of differentiation, as well as for pinpointing sources of environmental influence on behavior. She noted that the critical period concept does not explain the speech behavior that has been observed. She further noted that investigations of development in many species have indicated that it is not a critical period which imposes a temporal limit on development, but the unfolding of developmental processes themselves. It is Oyama’s view (1979, p. 88 ff) that the apparent superiority of early as opposed to late speech learning may support the existence of a broadly defined "sensitive" period. A sensitive period may be regarded as a period of heightened responsiveness bounded on both sides by periods of lesser responsiveness, or a period of "competence" for specific exchanges with the environment. According to Oyama, shifts in responsiveness to various environmental stimuli may co-occur with the "progressive elaboration of structures or schemata" which may affect how the organism engages the environment (see pp. 128).

Second, a number of studies have provided evidence that adults' ability to produce and perceive L2 sounds may, in certain circumstances, surpass that of children (see Flege, 1981). Winitz (1981), for example, found that native English adults were able to discriminate Chinese tones and obstruct consonants better than 8-year-olds. Snow and Hoefnagel-Höhle (1978) found that English-speaking adults produced Dutch sounds more authentically than children 8 to 10 years of age after a small amount of L2 experience.

Third, recent evidence has indicated that human cerebral functions and neuronal synaptic arrangements continue to develop long beyond the age of 12 years (see, e.g., Walsh & Diller, 1981). Moreover, there has been a great deal of controversy concerning the age at which cerebral lateralization (which is thought to affect L2 learning) reaches completion. Recent research has also suggested that lateralization for language functions may occur well before the age of 12 years, perhaps even by the age of five years or earlier (see e.g., Schnitzer, 1975). Further, it has never been clearly established that lateralization per se would impair language learning by either hemisphere (Oyama, 1982a).

The primary evidence offered by Lenneberg (1967) in support of the critical period hypothesis was that children, unlike adults, are capable of complete recovery from certain types of aphasia. However, evidence reviewed by Snow and Hoefnagel-Höhle (1982) fail to support this claim.

The critical period hypothesis leads to the expectation of a fairly abrupt difference between individuals of different ages relative to how authentically L2 is pronounced. This was not confirmed in a study of L2 pronunciation by Oyama (1982a). She found that degree of accent (as measured by global ratings) increased linearly as a function of the age (between 6 and 20 years of age) at which 60 Italian men first began learning English in the United States. There was no marked discontinuity indicating the end of a critical period at age 12 or any other age. Data reported by Tahta et al. (1981) did show a marked difference in global accent ratings between individuals aged 6 to 11 years and subjects older than 12 years of age when they first began learning L2 in the United Kingdom. However, the significance of the seeming nonlinearity was not tested statistically.

The critical period hypothesis also leads to the expectation that youngsters will be able to imitate L2 sounds better than older indi-
viduals. Cochrane and Sachs (1979) found that English-speaking 7-year-olds imitated Spanish words better than English-speaking adults, but Politzer and Weiss (1969) reported a positive correlation between age and the accuracy with which American school children imitated French vowels. Snow and Hoefnagel-Hohle (1977) investigated the imitation of Dutch words by 136 native English adults and children 5 to 17 years of age. They also found that the accuracy with which subjects imitated Dutch sounds was positively rather than negatively correlated with age.

Finally, and perhaps most importantly, many factors other than neurological maturation or organization might account for differences in L2 pronunciation between L2 learners of varying ages. Unfortunately, positing a critical period seems to have impeded the search for other explanations because it has created the impression in the mind of many researchers that the object of investigation had already been “explained.” One possible alternative hypothesis, for example, is that partial approximation to L2 phonetic norms is the result of incomplete learning rather than inability to learn. Some sounds appear to be inherently more difficult to learn than others, even by children learning their L1 (Macken & Ferguson, 1981). This suggests it will take any learner, child or adult, some finite amount of time to master the production of a new sound. The results of previous research may simply reflect adult phonetic learning at some midpoint of development, rather than at its eventual end point. Other alternative hypotheses are that children are more highly motivated to learn authentic L2 pronunciation than adults, receive more or better L2 input, or both. These alternatives to the critical period hypothesis, and others, will be discussed in the following sections.

L2 INPUT

The possibility exists that adult-child differences in pronunciation, when noted, may be due to differences in amount of experience with L2, even for subjects who may have lived for the same length of time in a L2-speaking environment. Burling (1981) noted that other adults impeded his progress in learning Swedish during a sabbatical year spent in Sweden. Burling’s colleagues often switched into English to promote a better exchange of ideas while his children were usually left to struggle in Swedish by their playmates, whose proficiency in English was probably less developed than that of most Swedish adults. Snow and Hoefnagel-Hohle (1978, p. 1116) made the observation that children learning Dutch naturalistically in the Netherlands received more L2 input than adults:

These observations were supported by answers given to a detailed language background and usage questionnaire administered by Cochrane (1977) to Japanese children and adults learning English in the United States. Compared to adults, the children tended to speak English with a larger number of people outside the home, and were obliged to use English in more social contexts than adults.

Second, the quality of L2 input may also differ for adult and child learners. Burling (1981), among others, noted that children tend to use language more often in reference to ongoing events and objects near at hand than adults. Adults are also more apt to discuss abstract concepts without a tangible referent. Asher (1981) observed that child L2 learners, unlike most adults, inhabit an “acquisition-enriched” environment in which much of the language addressed to them is immediately understandable from context (e.g., “Give daddy a kiss”; “Let’s wash our hands”). Language typically addressed to adults, on the other hand, tends to be less easily comprehensible (“Good morning”; “It’s a beautiful day”; “How are you?”).

This difference in quality of input may have two important effects. Children may understand a larger proportion of the L2 speech addressed to them than adults, leading to a quantitative difference in “intake.” L2 words may have a richer array of sensory associations for children than adults, making them easier to store and retrieve in speech processing. Asher and Price (1967) hypothesized that adult-child differences in pronunciation would disappear if L2 intake were truly equal for learners of different ages. In indirect support of this, they found that adults outperformed children in comprehending and executing commands spoken in a foreign language when exposed to physically demonstrated actions that were explained in L2 by the “caretaker” (instructor).

MOTIVATION AND AFFECTIVE FACTORS

Affective factors might also limit L2 pronunciation success. The extent to which the L2 learner approximates L2 phonetic norms may be
related to the extent to which they feel inclined (or obliged) to speak like a native speaker of L2. Macnamara (1973) suggested that part of the reason children seem to pronounce foreign languages better than adults is that they generally feel stronger pressure from their peers to conform to the phonetic norms of the surrounding speech community. Schumann (1976, 1978) asserted that affective factors are more important than age in determining success in L2 pronunciation, and suggested that adults' progress in L2 may be impeded by their fear of making mistakes or being ridiculed for communicating ineffectively in L2.

Schumann (1978) outlined a wide range of factors that might potentially affect L2 learning. Under “social factors” he included the socioeconomic status of the L2 learner vis-a-vis native speakers of L2. Individuals tend to emulate those with higher socioeconomic status, at least when using formal speech styles. Perhaps L2 learners who highly value L2 and the culture it is associated with will pronounce L2 more authentically than those who do not. Supporting this are the results of Hanlon (1971, cited by Cochrane, 1977), who found that children were more likely to imitate an individual’s accent if they identified positively with him or her. “Affective factors” included “language shock” and “culture shock.” Schumann speculated that anxiety induced by the many differences between L1 and L2, along with concomitant cultural differences, may impede progress. “Personality” factors included degree of extroversion, tolerance for ambiguity, sensitivity to rejection, and degree of self-esteem. These factors seemed likely to Schumann to affect the extent to which L2 learners seek out relevant L2 input and practice speaking L2.

Although interesting and plausible, Schumann himself noted that few of the factors he discussed have been shown empirically to predict L2 success. Moreover, few of them have been related directly to pronunciation, and none seem to have explicitly compared adult to child L2 learners. One study that did seem to offer support for the hypothesis that affective factors limit authenticity of L2 pronunciation is due to Guiora et al. (1972). They examined the effect of alcohol on L2 pronunciation, hypothesizing that alcohol intake would break down psychological inhibitions against pronouncing L2 authentically. Perceptual evaluations indicated that the imitation of Thai speech material by subjects who had ingested 1–1.5 oz of alcohol 10 min prior to a pronunciation test was better than that of subjects who had received a placebo. These subjects, in turn, were judged to have better pronunciation than those who had consumed 2–3 oz of alcohol.

Guiora (1972) argued that once an individual develops a sense of “language identity,” L1 pronunciation patterns become a manifestation of his or her personal identity. Once a language identity is established, pronouncing in new ways might be viewed as threatening to the learner’s “language ego.” Although other interpretations are possible, the Guiora, Beit-Hallahmi, Brannon, Dull, & Scovel (1972) results do suggest that some adults may be capable of pronouncing a foreign language better than they normally do.

In the view of many researchers, motivation plays an important role in determining how successfully a foreign language will be pronounced. Tahta et al. (1981) found that the extent to which learners used English in the home accounted for a significant 9% of variance in L2 pronunciation scores. When subjects aged 7 to 12 years were considered separately, it accounted for 26% of the variance. The authors speculated that this finding reflected a strong, positive identification with L2 and the culture it represented, rather than a difference in amount of input between subjects with relatively authentic and non-authentic pronunciations.

Gardner and Lambert (1972) described motivation as having an “integrative” and “instrumental” aspect. The learner who is integratively motivated wants to learn L2 in order to meet and communicate with native speakers of L2. The learner with an instrumental motivation, on the other hand, desires to learn L2 in order to achieve professional or social advancement. To the best of my knowledge, no research now available makes it clear how these different types of motivation might contribute to L2 pronunciation success.

Related to the issue of motivation is the effect of community attitudes toward L2 learning. L2 learning is generally not considered to be very important in the United States, as demonstrated by the relatively small number of students who study a foreign language beyond the minimum requirement of one or two years. This attitude stands in sharp contrast to that which prevails in certain other societies where the ability to communicate in several languages is an important determinant of status. Segalowitz and Gatbonton (1977, p. 86) noted that L2 learners attempt to pronounce L2 authentically when their community regards mastery of foreign languages as “prestigious and indicative of the speaker’s superior intelligence and level of education.” Hill (1970) observed that in cultures where L2 learning is a highly valued skill, even adults may manifest considerable success in L2 pronunciation.

One interesting possibility is that older L2 learners may have more reason to retain an L1 accent in L2 than younger learners. Although we outlined important negative effects of accent above, there are advantages as well. Certain varieties of speech have “overt” prestige, meaning they elicit positive evaluations for dimensions relating to status, competence, and power. However, native speakers of non-prestige social
dialects may assign “covert” prestige to their own form of speech for dimensions relating to qualities such as friendliness, trustworthiness, sense of humor, and so on.

If similar psychosocial phenomena influence L2 learners, whose accented speech might be regarded as a “non-prestige” dialect of L2, speaking with an accent may be perceived by the learner to be beneficial. More specifically, maintaining an accent may be regarded as a sign of loyalty to fellow L2 learners, especially those sharing the same native language (Ryan & Carranza, 1975). It might help the L2 learner preserve his or her ethnic identity in a community where L2 is the dominant language.

Supporting this hypothesis is the finding by Gatbonton (1975, cited by Segalowitz & Gatbonton, 1977) that the frequency with which French Canadians correctly produced the difficult /ɔ/ and /ø/ sounds in English words depended to some extent on how strongly they identified with the French-speaking population of Canada. Those subjects who were most nationalistic in their political views correctly produced these sounds less often than learners whose ethnic identification was less strong. Clearly, a “language loyalty” factor will be most relevant to L2 learners who inhabit a multilingual community where both L1 and L2 are spoken natively.

Another positive advantage of accent may be related to compliance with the social norms of the L2-speaking community. Social blunders which would be excused in an obvious foreigner might not be so readily tolerated in someone whose speech did not reveal them to be a non-native speaker. That is, a learner who speaks with a foreign accent may be excused for social transgressions in much the same way that children are excused. Neither foreigners nor children are acknowledged to be full participants in L2 society due to their relative lack of social knowledge. The L2 learner might attempt to avoid social censure by speaking with an accent that identifies her or him as not being a member of the community in the fullest sense.

STYLE AND SOCIAL MEANING

Variability is a characteristic of speech produced by native speakers as well as non-native speakers. Some variability in speech production derives from inherent limits on neuromotor control, and some from factors such as the talker’s mood. Still other variability is encoded into the speech signal, but unlike phonetically conditioned variability, sociolinguistically conditioned variability communicates social and attitudinal information in parallel with referential meaning. The most common object of sociolinguistic research is the “phonological variable,” an abstract sound unit comprising the range of possible pronunciation variants for a single sound. How a phonological variable is realized depends not only on linguistic factors such as degree of stress, but also on nonlinguistic factors such as age, ethnicity, gender, geographic origin, and social status. The choice of one variant over another (e.g., swimming versus swimmin’) can be predicted with some accuracy for members of various groups of native speakers in well-defined speaking situations.

Labov (1972) conceptualized speech styles as spanning a continuum from “careful” to “casual.” Formal contexts tend to evoke a careful style of speech in which the talker uses pronunciation variants considered to be typical of the “prestige” norm of the community. In casual speech, defined as “everyday speech used in informal situations” (Labov, 1972, p. 86), native speakers of “low prestige” dialects tend to use the “low prestige” variants of a phonological variable. Labov and others have found that the frequency of “low prestige” variants may increase across the following speech tasks: the production of minimal pair words; the reading of word lists; the reading of a story written in a colloquial style; speech produced spontaneously during a formal interview; and casual speech among friends. Native speakers often agree with one another concerning which of several variants of a phonological variable is “correct,” although they themselves may not characteristically produce that variant. Even children appear to have knowledge concerning which variant is appropriate in various speaking situations (Fischer, 1958).

Like some children who misarticulate sounds in L1, L2 learners sometimes appear to be aware of the “correct” pronunciation of an L2 sound even though they themselves may not produce it authentically. For example, many Japanese learners are painfully aware of the difficulty they have in distinguishing English /r/ and /l/. Many Arabs know they have a tendency to pronounce /p/ as /b/ (Flege, 1979); and German learners of English have heard many times about their tendency to devoice /b,d,g/ in word-final position. There is empirical support for the belief that L2 learners are consciously (or unconsciously) aware of some of their own L2 pronunciation errors. Neufeld (1979, 1980) found that learners who spoke French with an obvious English accent were able to distinguishing English-accented French from the French spoken by native speakers (see also Flege, 1984a). Brennan and Brennan (1981a)
found that Spanish speakers of English were able to quantify Spanish accent as accurately as native English speakers.

We presently know little concerning L2 learners' knowledge of sociolinguistically conditioned variation in L2. It seems likely that few L2 learners have the same tacit knowledge as L2 native speakers concerning which variants of a phonological variable are appropriate in various speaking situations, at least during the early stages of L2 learning. Even if L2 learners did possess such tacit knowledge, they might not possess the motor control needed to vary segmental articulation rapidly in conversational speech.

It is likely that L2 learners maintain learned L1 patterns of sociolinguistic variation in their production of L2, for the maintenance of L1 pronunciation patterns is common in L2 production. Several studies suggested that task-related differences in L2 pronunciation follow from patterns of sociolinguistic variation present in L1. Schmidt (1977) reported that certain Arabic words are realized with either a dental or interdental fricative (i.e., /s/ or /θ/), depending on whether classical or colloquial Arabic is being spoken. The interdental variant occurred frequently in the most formal of several Arabic speaking tasks (word list reading), but not at all in conversational speech. The dental variant, on the other hand, occurred more frequently in casual (including conversational) speech than in more formal speaking tasks. Schmidt noted a correlation between subjects' use of the interdental variant in "th" words in Arabic and English. In both languages it tended to occur most often in minimal pair word lists, less often in the reading of nonminimal pair word lists, and least frequently in the reading of paragraphs.

Beebe (1980) tested the hypothesis that L2 learners would use an English /r/ more often when reading lists of words than speaking spontaneously because reading permits subjects to direct more "attention" to their speech. She found, however, that her Thai subjects produced English /r/ correctly less often when reading a word list than speaking spontaneously because they tended to use a Thai-like (i.e., trilled) /r/ more often in the word-list reading task. Beebe concluded that the Thai subjects maintained a Thai reading convention when asked to read an English text. If so, transfer of an L1 sociolinguistic rule may have made it appear that their pronunciation had deteriorated when it was expected to improve.

An analogous observation was made by Major (1987), who reported that the authenticity with which Brazilians produce English /h/ may depend on speech style. In Portuguese, /h/ is realized as /x/ in certain contexts, which leads Brazilians to mispronounce /h/ in English words. However, Brazilian /x/ tends to be realized as /h/ in casual speech. If Brazilians transfer L1 sociolinguistic rules, it might appear that their realization of English /h/ deteriorates in formal compared to informal speaking contexts.

One other study showed that L2 pronunciation varied according to speaking task. L. Dickerson (1974, 1975) examined the ability of Japanese L2 learners to approximate the phonetic norms of English for a number of sounds. Pronunciation was more often—and more nearly—correct in word-list reading than in dialogue reading, and in dialogue reading than in conversational speech. Although the subjects' pronunciation of the sounds examined improved over the course of one year, the effect of speaking task remained. W. Dickerson (1976, 1977a, 1977b; L. Dickerson & W. Dickerson, 1977) hypothesized that the variability observed resulted from the same kind of tacit sociolinguistic knowledge native speakers use in L1 to vary systematically pronunciation. He asserted that the L2 learner's developing competence is best described by variable phonological rules specifying the probabilistic effect of speech styles and task, as well as phonetic context, on the selection of sound variants.

This hypothesis leads to the expectation that L2 learners develop an awareness of stylistically conditioned variation in L2. Supporting this is a study by Gatbonton (Segalowitz & Gatbonton, 1977), who created a context in which native English-speaking subjects were required to communicate in both casual and careful (i.e., formal) varieties of French and English. Rating scales were used to assess the subjects' evaluation of their own speech, as well as that of an absent interlocuter whose taped speech was used to condition variation in the subjects' speech. The subjects believed they sounded less intelligent and self-confident when they were induced to speak a casual variety of French than the careful variety taught in schools. The subjects also believed it was easier to express themselves and be understood in careful than casual French.

The English-speaking subjects attributed more negative personality characteristics to the interlocuter when he spoke a casual than careful form of French, perhaps as a result of being induced to attempt a style they did not fully control. Supporting this interpretation was the finding that subjects showed the opposite pattern of results in their native language, where they could presumably control stylistic variation to a greater extent. Also, native speakers of French attributed more positive personality characteristics to the interlocuter when he spoke casual than careful French.

Another indication that L2 learners are aware of stylistically controlled variation is the observation that they modify L2 speech production according to the identity of their interlocuter. It has been found
for example that, in speaking their L1, native speakers of British English vary the frequency with which word-final /t/ is realized as a glottal stop as a function of the social status of their interlocuter (Thekarar, Giles, & Cheshire, 1982). In “accommodating” their speech to that of an interlocuter, talkers may modify pause time, speaking rate, fundamental frequency, and segmental articulation so as to more nearly resemble their interlocuter (see Beebe & Zuengler, 1983). Beebe (1981) examined Chinese vowels produced by Chinese children who had learned Thai. The children used more vowels typical of Thai-accented Chinese when speaking to a Thai speaker of Chinese than to a native Chinese woman. Beebe felt the children’s change in vowel production was not simply the result of direct imitation, for the Thai woman was said to speak Chinese without a Thai accent.

In evaluating the hypothesis that variability in L2 speech production is rule-governed, it is important to note that the English sounds studied by Dickerson, Schmidt, and Beebe tended to be sounds that are difficult for L1 as well as L2 learners to pronounce (e.g., /r/, /l/, /s/, /θ/). Sociolinguistically or stylistically conditioned variability in L1 speech production presumably involves a choice between two or more sound variants that are controlled thoroughly by the speaker and can be selected without regard to motoric difficulty. Future L2 research examining patterns of alternation among variants that are of equal motoric difficulty is needed. Motoric difficulty might be assessed, in part, by the frequency with which sounds occur in babbling; the rate at which sounds are mastered by the majority of children during L1 acquisition; or by patterns of loss resulting from cerebral vascular accidents (see MacNeilage, 1980; MacNeilage, Hutchinson, & Lasater, 1981). If L2 learners are still found to resemble L2 native speakers, or if they manifest coherent patterns not seen in the speech of children learning L2 natively, it would support the hypothesis that L2 native speakers develop tacit knowledge of socially or stylistically conditioned patterns of variability in L2.

INDIVIDUAL DIFFERENCES

Individual differences are often found to exist before and after the acquisition of most motor skills (Fleishman, 1962). Not surprisingly, we see a wide range of authenticity in the L2 speech of individuals in most speech studies reporting individual subject data. Snow and Hoefnagel-Höhle (1982, p. 109) noted, for example, that native English speakers who were learning Dutch naturalistically over the course of a year:

Differed in . . . control of pronunciation. . . . One 7-year-old subject, for example, showed no improvement after session 2 (in imitating Dutch words), and his spontaneous pronunciation remained poorer than that of all the other (children) despite the fact that his scores on the other tests were relatively high. Conversely, a 12-year-old achieved almost perfect pronunciation very early, but continued to score poorly . . . on Sentence Judgment and Morphology, two tests which require the most metalinguistic ability.

Few learners pronounce L2 completely like native speakers, but all children pronounce their native language without accent. Thus, individual differences in pronunciation may be more obvious in early stages of L2 learning than in later stages of L1 acquisition because of the inevitable ceiling effects. Given this, it seems reasonable to suppose until proven otherwise that the same, largely unknown, individual factors which affect success in acquiring L1 will also affect L2 learning. Supporting this is a finding by van Balen (1980). He noted that Dutch high school students who received high scores on an English listening proficiency test also tended to receive relatively high scores on a comparable test in L1 using Dutch speech material.

Although most investigators recognize the existence of important differences in ability between individual L2 learners, we do not yet understand the basis for these differences, nor do we fully understand the relationship between pronunciation ability and other cognitive, perceptual, and motor skills. Snow and Hoefnagel-Höhle (1979) noted that most previous research relating to “language aptitude” has been aimed at delineating factors that might predict success in foreign languages learned through formal instruction rather than naturallyistically (the focus of the present chapter). The existing literature is therefore of limited interest because, in the present context, factors such as linguistic interest, associated memory, inductive learning ability, Grammatical Sensitivity are likely to measure components of educability as much as language learning ability.

One hypothesis that should be examined in future research is that individual differences in auditory processing lead to varying degrees of authenticity in L2 production. One early study (Pimsleur, Sundland, & McIntyre, 1964) revealed that differences in auditory processing ability, as measured by the Chinese Pitch Test and the Sound-Symbol Test, differentiated over- and under-achievers in the foreign language classroom. It is commonly accepted that musical ability, perhaps itself based on more general rhythmic and/or auditory processing abilities, is related to L2 pronunciation ability. Travis and Davis (1927; cited in Fay, 1979) reported that subjects scoring high on the Seashore
Measures of Musical Talent articulated their L1 (English) better than subjects receiving low scores. However, Thogmartin (1980; see also Tahta et al., 1981) reported that there was no relationship between the auditory properties of L2 sounds and children's success in repeating Chinese sentences during a 10-week class in Chinese.

Redmond (1977) proposed that L2 learners tend either to perceive L2 sounds in terms of the categories of L1 (a “code using” strategy) or to develop new central representations based on the actual auditory properties of L2 sounds (a “code forming” strategy). If such a dichotomy does distinguish groups of L2 learners, it might be expected to vary as a function of listening condition. For example, in early stages of learning when comprehension is limited, or in noisy listening conditions, learners may tend to favor a “code using” strategy. They might favor a “code forming” strategy, on the other hand, when processing easy-to-understand speech in optimal listening conditions. Redmond (1977) noted another dichotomy distinguishing individuals: the extent to which individuals centrally fuse auditory information presented to the two ears. He speculated that “fusers” may show a greater tendency to overlook the allophonic variations which distinguish sounds in L1, as well as the subcategorical phonetic differences which may distinguish similar sounds in L1 and L2.

Future research should attempt to determine what relationship exists (if any) between learners' ability to imitate sounds accurately and authenticity of L2 pronunciation. The ability to pronounce new as well as similar L2 sounds ultimately depends on the learner's ability to accurately assess the auditory properties of L2 sounds, and to translate this information into motor commands. Locke (1969) examined the ability of preschool children to imitate the sounds in a Swedish disyllable. Children who pronounced their L1 (English) well were judged to imitate Swedish sounds more accurately than children receiving lower scores on a standard English articulation test.

Lerea and LaPorta (1971) examined the ability of native English-speaking adults to imitate the unfamiliar voiceless uvular fricative /x/ in a list of Hebrew words. The first group consisted of “compound” bilinguals who began their study of a foreign language (one without /x/) in high school or college. The second group consisted of “coordinate” bilinguals who began to study an L2 at about the age of five years. The third group was made up of monolinguals who had apparently never studied a foreign language. After nine presentations of word lists, the “compound” bilinguals were judged to produce /x/ authentically in about twice as many words as the monolinguals, with the coordinate bilinguals intermediate in accuracy. The authors suggested that the compound bilinguals' recent experience learning a foreign language via the “audio lingual” method may have improved their ability to imitate unfamiliar L2 sounds. It is not clear why the monolinguals performed poorly in imitating /x/. Their choice not to study a foreign language might indicate self-awareness of a relatively poor ability or fear of pronouncing foreign languages. But if they had simply never had the opportunity to study a foreign language, it might indicate that lack of experience learning new sounds diminishes the ability to do so.

Locke (1970) examined the ability of English-speaking children 4 to 11 and 8 to 12 years of age to imitate two consonant and one vowel sound not found in English. Native speakers assessed authenticity using a 15-point rating scale. Most improvement for all three sounds occurred between the first and second imitation trials, with little improvement from trials 3-10. Earlier studies had shown continued improvement over the course of 20 trials. Locke speculated that the seeming discrepancy between studies may have been due to a procedural difference (the presence versus absence of feedback) or motivational differences. However, Locke's results were replicated in a second experiment, in which 68 four-year-olds imitated a Swedish syllable 20 times, this time with feedback and reinforcement. For each sound in the syllable, most (40% to 83%) improvement occurred from trial one to trial two.

Imitation data are difficult to interpret because they are shaped by many factors, including perception of an auditory stimulus (which may entail identification of its structural characteristics); storage and retrieval from short-term memory of information derived from auditory processing; and the translation of stored information into motor commands to the speech articulators. Successful imitation may skip a stage of phonetic processing altogether, as demonstrated by the ability of the mynah bird to mimic accurately the acoustic structure of human speech sounds using a vastly different oral mechanism (Marler & Mundinger, 1971).

The auditory monitoring of imitated sounds may reveal to an individual who is talking the extent to which he or she has matched a sound by providing an “error” message. One possible interpretation of Locke's results is that the error message generated after imitation on the first trial triggered a search for a more adequate articulatory strategy that exhausted the subjects' capacity to further narrow the perceived error (at least immediately). Further improvements might have been meager because the subjects settled on a known motor response (e.g., substituting an L1 sound for the L2 target sound). It is possible that further changes in production would have depended on changes in central phonetic representations, and that such changes occur slowly as the result of monitoring many attempts at producing a sound.

It is unclear at present to what extent L2 learners differ in their
ability to imitate L2 sounds. In a preliminary study, Pike (1959) had some 200 phonetic students from a variety of L1 backgrounds imitate their teacher's pronunciation of speech sounds not found in English (e.g., voiceless vowels and nasals, nasalized vowels, a voiceless velar fricative). Students were also asked to perform a number of speech-related vocal gestures based on verbal instructions (e.g., “Prolong the /s/ in slow.”; “Make the /l/ in allow voiceless.”). Pike reported a significant correlation between scores on this “phonetic ability” test and the grade assigned to each student after 45 hours of instruction in phonetics.

These results must be interpreted cautiously. Students were assigned to different sections of the phonetics class on the basis of their test scores, which may in itself have influenced the final course grade and accounted for the positive correlations observed. Nonetheless, these results suggest the possibility that ability to imitate may predict the L2 learner’s ultimate success in pronouncing an L2. This conclusion is further supported by a study (Suter, 1976) which reported a significant positive correlation between rating-scale judgments of the English accent of 61 native speakers of Japanese, Thai, Farsi, Arabic, and the performance of these L2 learners on an imitation test similar to Pike’s. Four other variables out of the 20 examined were also significantly correlated with overall English accent (i.e., the speaker’s concern for pronunciation, the percentage of time English was used outside the home, the age at which English was learned, and the number of years spent in the United States). Unfortunately, the author did not test the relative strength of these variables, nor perform partial correlations in order to explore possible interrelationships among variables.

It would be useful to test the ability of individuals just beginning the process of L2 learning to imitate sounds found in L2 as well as other foreign languages. Subjects might be required to imitate the same speech material at varying periods of delay (e.g., no delay or delays of 0.2 sec, 0.6 sec, 1 sec, and 5 sec) after aural presentation to explore the effect of phonetic coding on imitation. It would also be useful to examine whether variables such as concern for pronunciation, age of first exposure to an L2, amount of time in an L2 environment, or age are related to imitation ability.

LISTENER EFFECTS

So far we have considered only factors which might affect the learners’ production of L2 sounds. However, the extent to which the speech of the L2 learner is perceived to be authentic also depends on the listener. It is well known, for example, that listeners perceive a sound to be present when it has been omitted or obliterated by noise (the “phoneme restoration effect”). Words are generally identified better in the context in which they were originally produced than when edited out and presented in isolation, especially when that context contains relevant semantic information. When the acoustic specification of a word is ambiguous, listeners tend to hear a word conforming to the expectations generated by the preceding speech material (Bond & Garnes, 1980).

Gladstein, Membrino, and Flégé (1983) found that listeners’ use of “higher order” information may obscure an objective assessment of the speech sounds produced by L2 learners. English-speaking listeners without training in phonetics wrote down the final word of sentences produced by six Vietnamese learners of English. The identity of half of utterance-final words was predictable from preceding context (e.g., “He put the food in the shopping bag”) and unpredictable in the other half (e.g., “Please repeat the word bag”). Listeners correctly identified the voicing characteristic of word-final stops (/b,d,g/ versus /p,t,k/) about three times more often in the predictable than unpredictable sentences. As a result of using semantic context to recognize words, the untrained listeners failed to perceive many instances of final stop devoicing that were perceptually evident to trained listeners who had transcribed the same words. This finding suggested the possibility that the L2 pronunciation of an individual might appear to improve without an actual change in authenticity. That is, as the L2 learner produces more semantically complex sentences, listeners may simply become better able to identify correctly the words they produce by exploiting semantic information in addition to acoustic and visual information.

PHONOLOGICAL AND PHONETIC FACTORS

A TWO-LEVEL MODEL OF PRODUCTION

Speech production has typically been viewed as a two-stage process in which relatively abstract sound units are first selected at higher levels of a message generating system, then relayed to a lower-level system for transformation into a code suitable for the generation and control of articulatory movements. An important implicit assumption of most research in L1 and L2 speech learning (see, e.g., Flégé & Port, 1981; Locke, 1983b) is that nonauthentic pronunciation arises either from incorrect selection of a sound unit (at a “phonological” level of organization) or from incorrect implementation or realization (at a “phonetic” level).

An important source of “selection” errors seems to be an incorrect lexicalization of L2 words and morphs. For example, Arabs’ lexical
representation for English “pat” might be /baet/ because the initial sound is perceived as /b/ due to the absence of a /p/ phoneme in Arabic (Flege, 1979). Similarly, the inability of many Japanese learners to distinguish /r/ and /l/ seems to stem from uncertainty concerning which English words contain /r/ and which contain /l/.

The listener might perceive a segmental substitution or distortion as the result of incorrect phonetic implementation or realization. For example, an Arab learner might correctly lexicalize “pat” as /pæt/, but be heard to have said “bat” because of an insufficient knowledge concerning how English /p/ is implemented. Similarly, the Japanese learner’s attempts at /l/ might be heard at times as /t/ because he or she has not yet developed the ability to phonetically implement the distinction between /r/ and /l/.

Such a two-level model explicitly or implicitly underlies much recent work in normal and disordered speech motor control. Many researchers agree that speech sound production is guided by targets that are “feature specified” or “goal oriented” (Perkell, 1980, p. 347). In linguistically-oriented models of speech production, the targets are phonemes: abstract, timeless units conceptualized as discrete and nonoverlapping units of mental representation. The two primary reasons for believing that it is phonemes rather than context-sensitive allophones which represent the targets to be learned is that human languages all share the same characteristic of recombining a small number of sound types to produce a much larger number of lexical items. More importantly, listeners perceive the same sound when a phoneme is realized differently in various phonetic contexts and syllable positions.

MacNeilage (MacNeilage et al. 1981; MacNeilage, 1980) presented several arguments for the distinction between phonological and phonetic levels of organization. In his view, the importance of a “pre-motor” (or phonemic) level of organization is supported by the existence of alphabetic writing systems, and by the permutation of sounds observed in slips of the tongue. This last argument is compelling since it appears that in “spoonerisms” (e.g., “tut your cape” for “cut your tape”), phoneme units are sequentially permuted prior to the formulation of the motor commands needed for phonetic implementation. That is, sounds are produced with the phonetic details (e.g., the presence or absence of aspiration in /t/) appropriate for the phoneme in its new (i.e., permuted) location in the speech stream. Spoonerisms do not appear to be motoric (or phonetic) in nature since they do not generally involve the replacement of relatively difficult (or “marked”) sounds by relatively easy-to-produce (or “unmarked”) sounds (see Eckman, 1977, and Rutherford, 1982, for discussions of markedness as it relates to L2 pronunciation).

The validity of a two-level model for studies of L2 learning is supported by the finding (Monnot, 1972) that native speakers of American English can easily learn to pronounce a Spanish /r/, which is realized as a single tap in Spanish, by substituting the flap which occurs in American English words like “ladder” and “latter” instead of attempting directly to produce Spanish /r/. That is, native English speakers may succeed better in producing Spanish /r/ by using a sound which occurs on the phonetic surface of English (i.e., a flap), even though it belongs to a phoneme other than /r/, than by modifying their implementation of English /r/.

A motoric (or phonetic) level of organization is supported by several kinds of evidence. One is the ubiquitous presence of allophonic variation due to phonetic context or position. For example, /e/ can be produced with similar acoustic properties either with the jaw clenched or free to assume its normal position. To take another example, voicelessness may be achieved during the closure for /p/ by either tightly constricting the vocal folds (as in “scrapbook”) or abducting them (as in “upon”). Cross-language differences also provide compelling evidence that talkers must acquire highly detailed knowledge of how sounds are to be realized phonetically (see Flege, 1984a).

Research on speech disorders following brain damage suggests that impairments may stem from disruption of the neuromuscular organization needed for the articulation of sounds. They may also arise from disturbances affecting the abstract linguistic system underlying the generation of speech and its comprehension, or from disturbances affecting the abstract coding or sequencing of sounds (Lesser, 1978). MacNeilage et al. (1981) marshaled evidence that brain damage can separately affect phonological and phonetic representations used to guide the production of speech (see also Schouten, 1975). They noted that nonfluent anterior (i.e., Broca’s) aphasics are often viewed as having a speech disorder that affects phonetic implementation. Sounds thought to be “difficult” (i.e., those occurring with relatively low frequency in infants’ babbling and various human languages) tend to be replaced by “easy” sounds in the speech of such apasics. Fluent posterior (i.e., Wernicke’s and Conduction) aphasics, on the other hand, tend to implement sounds correctly but may select or sequence phonemes incorrectly. As a result, their correctly articulated words are often meaningless. For nonfluent aphasics, there is a bias towards the use of motorically “easy” sounds, but this bias is much less evident in their speech than in the speech of fluent aphasics.

A two-level model also underlies generative phonology (e.g., Chomsky & Halle, 1968), which posits that learned, language-specific phonological rules may transform phonemic representations of speech sounds before
commands to the speech articulators are formulated. Learned phonetic realization rules are used to transform the binary distinctive feature specification of phonemes into more detailed articulatory specifications used to guide articulatory movements in time and space.

Research (e.g., Flege, 1980; Flege & Port, 1981) has suggested that an important necessity of phonetic (as opposed to phonological) learning is to acquire a knowledge of how phonemes are to be transformed into the dynamic gestures needed to realize L2 speech sounds (see Perkell, 1980; Stevens & Perkell, 1977). However, it should be noted that the distinction between phonetic and phonological levels of representation may be a more nearly quantitative than qualitative distinction. Even systematic phonetic representations are abstract since they “average over” many different phonetic contexts, speaking rates, degrees of stress and so on.

**PHONOLOGICAL RULES**

Arguing for the importance of phonological rules in L2 learning is the observation that L1 phonological rules may be maintained in L2 production with little or no modification. For example, American English is said to have a rule which changes /t/ and /d/ into flaps in post-stressed, intervocalic position. When the flapping rule applies, as it usually does in American English speech, “latter” and “ladder” become homophonous. Native English speakers know that the words “latter” and “ladder” contain a /t/ and /d/, respectively, as evidenced by their spelling and careful pronunciation of these words. Moulton (1962) observed that Americans inappropriately flap /t/ and /d/ in German words, despite the fact that monolingual native speakers of standard German do not.

Flege and Davidian (1985) examined the production of word-final stops by 12 native speakers each of American English, Polish, Chinese, and Mexican Spanish. Even in ideal speaking conditions (i.e., the picture-cued production of familiar minimal-pair words) the Mexicans realized /b,d,g/ (but not /p,t,k/) as homorganic fricatives in about 40% of instances. The authors concluded that a Spanish phonological rule had been maintained in English L2 production because the observed substitution pattern conformed to the general specification of a Spanish phonological rule (Harris, 1969). Moreover, fricativization of final stops is not common in English L1 acquisition, and was not evident in the speech of the Chinese or Polish subjects.

L2 learners may produce L2 with phonological rules found in neither L1 or L2. Eckman (1981) observed that two native Spanish speakers (Colombians) fricativized stops in English words. However, he concluded that his subjects were using an “interlanguage” phonological rule rather than a rule found in the phonology of either monolingual speakers of Spanish or English speakers. Eckman (1981) also observed that two “advanced” Chinese speakers of English inserted a centralized vowel at the end of CVC English words (e.g., /teg/ for “tag”). He hypothesized that these Chinese learners had developed a rule of post-stop vowel insertion (said to be nonexistent in human languages) in response to a phonotactic constraint of Chinese phonology which prohibits stops in word-final position.

However, several interpretations other than “rule invention” are possible. Major (1987) hypothesized that post-stop vowel insertion by Brazilian speakers of English resulted from a “universal” phonological process similar to that which affects the speech of certain children. Children acquiring English as their L1 often appear to simplify the production of CVC words by producing them as CVCC words. It is also possible that the vowels inserted post-vocally were nonfluently hesitated pauses. Hesitation pauses can be observed in the conversational speech of native English speakers (Shockey, 1974). However, there is no empirical evidence that permits us to determine whether they occur at a less frequent rate than was observed in the English spoken by Eckman’s native Spanish talkers.

Little evidence is now available concerning L2 learners’ ability to stop using an L1 phonological rule not appropriate for L2 production. It would be interesting to learn, for example, whether experienced native speakers of German stop flapping /t/ and /d/ in German words. The results reported by Flege and Davidian (1985) suggested that many adult L2 learners persist in using an inappropriate L1 rule in L2. The native Spanish speakers in that study continued to fricativize English /b,d,g/ after having lived 1 to 17 years in the United States.

An understanding of the role of phonological rules is of potentially great importance, for the continued application of L1 phonological rules might actually preclude phonetic learning. Consider, for example, the individual who is a native speaker of a Spanish dialect in which /s/ is realized as /h/ in word-final position. Such an individual might fail to perceive the substitution of /h/ for /s/ as the result of “backwards derivation,” which would cause a surface /h/ to be heard as underlying /s/. Backwards derivation appears to be needed to explain why adults, for example, tend to hear the child’s “wabbit” as “rabbit” (see Lovins, 1976). Broselow, Hurtig, & Ringen (1983) provided evidence that backwards derivation may cause native English L2 learners to misidentify tones in Chinese. If backwards derivation prevents the learner from noting L2 pronunciation errors, he or she may never set out to acquire the motor skills needed to produce correctly the errored sound.
Phonological rules may need to be learned in L2 just as new sounds need to be learned. Unfortunately, we have only limited information concerning the learning of new phonological rules in L2. A study by Port and Mitleb (1983) provided evidence that L2 phonological rules can be learned. Jordanian subjects who had never lived in an English-speaking environment seldom flapped /t/ and /d/ in English words, whereas subjects who had lived in the United States did. In contrast, Flege (1979) found that Saudi Arabian learners of English who had lived in the United States for as long or longer than the Jordanians in the Port & Mitleb (1983) study seldom flapped English /t/ and /d/.

The basis for the seeming difference between the Jordanians and Saudis is not clear, but it is unlikely to be motoric in nature. Both Saudis and Jordanians realize /r/ as flaps in their native dialect of Arabic. Perhaps the difference arose because the Saudis had learned to flap but, like certain native speakers of English, they regarded it as inappropriate to do so when reading (see also Beebe, 1980).

We lack knowledge at present concerning the relative modifiability of phonological rules and phonetic realization rules. Port and Mitleb (1983) argued that phonological rules are “formulated in the nervous system” in such a way that they are more easily modifiable than phonetic implementation or realization rules. Broselow (in press, b) suggested that “low level” phonetic rules are maintained to a greater extent in L2 production than phonological rules. By “low level” Broselow meant allophonic rules which have an obvious and exceptionless conditioning context, or surface phonotactic constraints. According to her definition, phonological rules admit exceptions and may be morphologically or syntactically conditioned. Major (1987) suggested that it may be easier for learners to overcome the substitution of one type of sound for another (e.g., use of Portuguese /r/ for English /r/) than the inappropriate application of a phonetically conditioned and exceptionless phonological rule (“process,” in his terms). He speculated that L2 learners are more “conscious” of the former than the latter.

A great deal of further research is needed to understand the extent to which phonological rules are maintained, learned, or lost in L2 learning. Future research is also needed to determine whether it is easier to learn suprasegmental or segmental rules, and whether it is easier to learn/modify phonetic realization rules or phonological rules. In addition, the hypothesis (Major, 1987) that learners are less sensitive to within-category acoustic variation than to cross-language differences in the implementation of a category (e.g., differences between the /r/ of English and Portuguese) needs to be tested.

**PHONOLOGICAL AND PHONETIC FEATURES**

An important question for speech research is “What are the units of speech which are learned?” Are whole word units learned, as sometimes appears to be the case in early stages of L1 acquisition? It is part or all of a syllable that is encoded and represented centrally? Or does each context-sensitive allophone need to be learned? Evidence from a variety of sources points to the importance of each unit, but since most L2 research has focused on segmental articulation, the discussion to follow will center on the properties of speech sounds produced by the L2 learner. Specifically, we will address the question of whether speech sounds are learned as undifferentiated wholes, or whether specific properties (or “features”) of sounds are learned separately and then combined into speech sounds.

Prague School phonologists (e.g., Trubetzkoy, 1939/1969) viewed a phoneme as a co-occurring bundle of linguistically “distinctive” and “redundant” features pertaining to specific aspects of sound production. One benefit of describing phonemes as a “bundle of features” is that it makes possible the non-instrumental comparison of sounds occurring in different languages, as well as a comparison of sounds produced by various groups of individuals such as those with disordered or developing speech (Lesser, 1978). An important assumption of this approach is that phonemes in the sound systems being compared can be accurately described in terms of the same set of features.

The features used in phonological analyses have generally been based on traditional articulatory phonetic taxonomies, or defined in terms of measurable acoustic properties (e.g., “grave,” “vocalic”). However, phonologists have proposed and used features primarily with the aim of taxonomizing contrastive sound units rather than describing the production, perception, or learning of sounds. Phoneticians, on the other hand, have focused more closely on properties associated with the physical production and auditory perception of sounds. Their ultimate aim is to understand the control parameters used to signal meaning differences in a vocal communication system. An important question in experimental phonetics is “What dynamically changing vocal tract configurations do talkers aim to achieve when they select particular sounds for production?”

Research in phonetics has tended to emphasize either the auditory-acoustic or the physiological properties of sounds. MacNeilage (1970) proposed that speakers aim to achieve invariant configurations of the oral cavity that are defined spatially and temporally. Ladefoged (1971) argued that while some features are based on articulatory properties...
of sounds, others are defined auditorily. The hypothesis of Nooteboom (Nooteboom, 1973; Nooteboom & Doodeman, 1980) is that talkers attempt to reproduce through their speech gestures a detailed auditory-derived code derived from previous experience with the many particular instantiations of a sound. In this view, the target for a vowel might be idealized formant values (Lindblom, 1963). Another view is that talkers attempt to regulate just those aspects of vocal tract shape that have auditorily important acoustic consequences. In support of this, Lindblom, Lubker, and Gay (1979) observed that when a talker produces a vowel many times, the smallest amount of token-to-token variability in tongue-palate distances occurs at that place along the vocal tract where variation would result in the greatest acoustic change (usually the place of narrowest constriction).

Still another view is that talkers attempt to recreate known patterns of sensation in the oral cavity. Perkell (1980; Stevens & Perkell, 1977) proposed that the talker attempts to achieve a pattern of orosensory stimulation associated with the auditory properties of sounds. An orosensory feature for a high vowel such as /i/, for example, might be the sensation of contact between the lateral margins of the tongue and the maxillary teeth. An orosensory feature that might be used in place of the distinctive feature “continuant” to specify an obstruent sound is the sensation of relatively high intraoral air pressure, signaled by pressure-sensitive mechanoreceptors in the upper airways and glottis.

An assumption made by some researchers is that speech learning in L1 acquisition can be described in terms of the acquisition of distinctive features (Jakobson, 1968). Misarticulating children have been characterized as being unable to control the production of distinctive features such as “strident,” “continuant” or “voice.” The principal evidence in support of this view is the finding that children may evidence the same kind of difficulty producing a natural class of sounds sharing the same phonological feature (McReynolds & Bennett, 1972). In keeping with this view, Koo (1972, p. 151) suggested that L2 learners need not actually learn to pronounce L2 sounds differing from L1 sounds, but instead must learn to “extend or reduce” native language sounds by “adding or dropping a feature . . . to or from the native sound he already has.”

Walsh (1974) questioned the utility of distinctive features in describing the process of speech acquisition. He argued that distinctive features often fail to provide meaningful insights into speech acquisition because they tend to focus on just those “selected properties” of sounds which are most useful for describing phonological contrasts between pairs of sounds. He noted that distinctive features often “encompass too great an area of phonetic space “to be maximally useful for speech analysis and treatment. For example, both /t/ versus /d/ and /t/ versus /s/ differ according to a single binary distinctive feature (“voice” or “continuancy”). However, both sound pairs differ continuously along a much larger number of physical parameters which include the place and force of tongue-palate contact, the portion of the tongue making contact, tongue shape, and the duration of contact. Walsh (1974) argued that differences between developing and mature speech should be described in terms of the details of articulation used to produce a sound, including redundant (i.e., nondistinctive) and language-specific details of phonetic implementation.

Generative phonology (e.g., Chomsky & Halle, 1968) assigns responsibility for articulatory details to a set of language-specific “phonetic implementation” rules (Chomsky & Halle, 1968) rather than abandoning the claim that human speech sounds are best described by a small set of universal distinctive features. This represents the implicit hypothesis that phonetic features must be learned in addition to, or perhaps instead of, distinctive (i.e., phonological) features. If so, the English-learning child who substitutes /s/ for /θ/ must learn to produce a new sound, one in which the tongue tip and blade contacts the alveolar ridge instead of the upper incisors, rather than learning to produce the distinctive feature “strident” (or, even more abstractly, learning to contrast “strident” with “nonstrident” sounds).

One assumption commonly made about both phonetic and distinctive features is that they represent independently controllable aspects of articulation. An important prediction following from this view is that features can be recombined to form new or different sounds. Grillner (1981, page 218 ff.) speculated that learned movement patterns, including those used to produce speech sounds, result from the combination of neural motor “subunits” into larger motor programs that are integrated at “higher centers,” and that different sounds result from different combinations of neuromotor “subunits.”

The property of “commutability” should, in principle, apply as much if not more to distinctive as to phonetic features. This is because distinctive features are linked to articulation only through phonetic rules of implementation. For example, Arabic has a labial stop /b/ and the voiceless stops /t/ and /k/. If distinctive features can be recombined freely, then Arabic speakers of English should be able to produce /p/ by recombining features whose implementation they already control. Flege (1979, 1980; Flege & Port, 1981) found this was not so. Arab learners produced the same temporal difference in closure duration between English /p-b/ as between /t-d/ and /k-g/, demonstrating tacit knowledge that the same distinction applied to all three stop voicing pairs. However, although their /t/ and /k/ were clearly recognizable,
they often produced /p/ with voicing so that it was frequently heard as /b/. Thus they appeared unable to combine reliably the voicelessness of /t/ and /k/ with the labiality of /b/.

Two other studies suggested that features are not freely recombinable. Fischer-Jorgensen (1968) examined stop production by a Danish-French bilingual. Her subject produced stops in Danish and French with the stop closure durations appropriate for voiced and voiceless stops in those two languages. However, the bilingual subject produced /b,d,g/ without voicing, something appropriate for voiced stops in Danish but not French. Suomi (1976) examined the English stops produced by Finnish learners. They more nearly approximated the temporal norms of English for stop closure duration than the difference in voicing which often distinguishes English /b,d,g/ and /p,t,k/.

NEW VERSUS SIMILAR L2 SOUNDS

Certain sounds found in L1 and L2 are realized in a seemingly identical fashion, so L1 and L2 might be said to possess the “same” sound. In other instances, however, a sound in L2 bears sufficient resemblance to an L1 sound that it is judged as belonging to the same phonetic category, yet is realized in a systematically different way from its counterpart in L1. We will refer to L2 sounds with a close yet acoustically different counterpart in L2 as a “similar” sound. A good example of similar sounds are the /t/ sounds found in French and English. The French and English varieties of /t/ are transcribed using the same IPA symbol, yet manifest significant acoustic differences that can be detected even by untrained listeners (Flege, 1984a; see also Flege & Hammond, 1982; Elsendoorn, 1983a, 1983b, 1984).

Still other sounds in L2 do not bear sufficient resemblance to any sound in L1 that they are likely to be regarded by the experienced L2 learner as belonging to an already familiar L1 category. We will refer to such sounds as “new” L2 sounds, although it should be apparent that any sound used productively in a human language, no matter how exotic, will bear some degree of articulatory and acoustic similarity to sounds found in the learner’s L1.

An assumption underlying most segmentally oriented studies of L2 production (e.g., Lado, 1957, Wode, 1977, 1978) is that learners tend to produce L2 words with L1 sounds. The seemingly frequent substitution of L1 for L2 sounds by adult learners of L2 may be due in part to the fact they, unlike children acquiring L1, often do not avoid attempting sounds they can not yet control. Adult L2 learners, like some children acquiring L1, usually adopt the procrustean solution of using a sound they already control in place of one they can not yet produce authentically.

One conclusion to be drawn from past research is that the overall phonetic similarity of sounds triggers the substitution of an L1 for an L2 sound. Despite the fact that a single mechanism (i.e., maintenance) leads to the substitution of an L1 sound for an identical, similar, or new L2 sound, the effect on authenticity may be quite different. Maintaining the use of a previously learned L1 sound goes unnoticed when an identical L1 sound is used in L2. This is referred to a positive transfer. The use of an L1 sound for a similar L2 sound (negative transfer), on the other hand, may be regarded as a distortion or substitution, depending on the degree of acoustic similarity between the L1 sound and its L2 counterpart.

Speech-language pathologists have long debated whether it is easier to help a child modify the production of a distorted sound or to teach the child to produce a new sound (Milisen, 1954). The same question has been raised in regard to L2 speech learning. Is it easier to learn the pronunciation of new L2 sounds which are not identified with L1 sounds than to learn to modify production of L1 sounds that merely resemble a target L2 sound? Valdrnan (1976, p. 38) speculated that it is easier for adults to learn new L2 sounds because they require “complete sets of new articulatory habits” and will tend not to evoke inappropriate articulatory “habits” established for the production of L1 sounds (see also Koutsoudas & Koutsoudas, 1962).

The results of several studies have suggested that learners produce similar L2 sounds more authentically than new L2 sounds. Brière (1966, 1968) taught French, Arabic, and Vietnamese sounds that were either new or similar for native English-speaking subjects. The subjects’ attempts were transcribed by native speakers of those languages. The test sounds with “close” phonetic equivalents in English (i.e., the similar sounds) were generally judged to have been produced more authentically than the new sounds which did not have a close counterpart in L1. There were, however, large differences between subjects in ability to produce some of the new sounds. For example, the voiceless uvular fricative /x/ was learned faster and better than its voiced cognate /y/ even though neither sound occurs in English. Politzer and Weiss (1969) found that English-speaking children 6 to 14 years of age succeeded somewhat better in producing the similar French vowels /ʒ/, /œ/, and /œ/ without the diphthongization characteristic of their English counterparts than in producing new French front rounded and nasal vowels.

As mentioned earlier, Flege and Port (1981) found that Arab learners produced the new sound /p/ in English less authentically and adequately than English sounds with counterparts in Arabic (i.e., /t/ and /k/).
However, one study indicated a clearly superior production of new compared to similar sounds. Flege (1986) examined the production of the new French vowel /y/ and the similar French vowel /u/ by three groups of native English speakers differing in L2 experience. The least experienced L2 learners failed to match French monolinguals in producing either /u/ or /y/. The mean second formant frequency (F2) of their /y/ was too low, and the F2 of their /u/ too high by French phonetic standards. Talkers in the two most experienced groups did not differ from French monolinguals in producing the new vowel /y/. However, even subjects in the group that was most experienced differed significantly from French monolinguals in producing the similar vowel /u/.

The apparent discrepancy between the results of Flege (see also Flege & Hillenbrand, 1984) and the results of the studies mentioned earlier may have been due to procedural differences or to the kind of subjects studied. For example, the transcriptions of Walz (1979) indicated that undergraduates in their first semester of French usually substituted an /u/-quality vowel for French /y/. This suggested that beginning L2 learners may produce the new French vowel /y/ less adequately than the similar French vowel /u/. It is possible that similar sounds are produced more authentically than new sounds in early stages of L2 learning (Brière, 1966) because the learner simply uses a familiar L1 sound. In other words, the similar sound gets a “free ride” initially. At later stages, however, new sounds may be produced more authentically than similar sounds because previous phonetic experience does not exert an influence.

Another possibility is that no generalization can be drawn for all new and similar sounds. F. Johansson (1973) transcribed the Swedish sounds produced in an imitation task by 180 subjects representing nine different L1 backgrounds (Serbo-Croatian, Czech, English, Danish, Finnish, Greek, Hungarian, Polish, and Portuguese). The range of errored productions decreased over time. The “hierarchy of difficulty” for Swedish speech sounds that Johansson constructed indicated that several of the sounds produced most authentically by subjects in all nine subject groups were those identified by Jakobson (1968) as among the first, and therefore presumably “easiest,” sounds acquired in early childhood. However, the most difficult sounds for subjects of all language backgrounds included both new and similar L2 sounds.

**PERCEIVING L2 SPEECH SOUNDS**

In this section we will explore the perception of L2 sounds by native and non-native speakers. Research often shows a close alignment be-

tween specific aspects of sound production and the perception of a sound by mature native speakers. Since sounds are often realized differently across languages, listeners may need specific experience to identify and discriminate optimally the speech sounds of L1. And since relatively few sounds encountered in L2 are produced exactly like sounds in L1, it is likely that L2 learners will perceive at least some L2 sounds differently from native speakers of L2.

The most direct and obvious confirmation of the expected perceptual differences between native and non-native speakers is the observation that L2 learners tend to use L1 category labels for L2 sounds. For example, Scholes (1967) found that subjects labeled a matrix of synthetic vowels using the vowel categories of their native language (see also Lotz, Abramson, Gerstman, Ingemann, & Nemser, 1960, for results pertaining to stop consonants). Schouten (1975) presented synthetic vowels to native speakers of Dutch and English. Subjects labeled stimuli using the vowel categories of their native language, and Dutch subjects who spoke English also attempted to label stimuli using the vowel categories of English. Many of the Dutch subjects labelled as /e/ stimuli heard as /i/ by English subjects, and labelled as /æ/ stimuli heard as /e/ by native English subjects (apparently due to the absence of an /æ/ category in Dutch).

Speech sound categorization is the endpoint of a complex perceptual process. The “incorrect” identification of an L2 sound may provide evidence for the maintenance in L2 of patterns of speech processing established for L1. For example, Broselow et al. (1983) examined the identification of Chinese tones. The English L2 learners made few identification errors, except for tones associated with utterance-final syllables, where they identified a high falling tone (tone 4) as a high level tone (tone 1). It appeared that the native English subjects, accustomed as they were to hearing a high falling Fo contour associated with a stressed, utterance-final syllable, judged the high falling (tone 4) Fo pattern as a high level (tone 1) contour that had been affected by utterance position. That is, they may have tried to “recover” the underlying linguistic content of the Fo contour through backwards derivation (see Lovins, 1976).

**FACTORS AFFECTING SEGMENTAL PERCEPTION**

General statements regarding the ability of L2 learners to “perceive” an L2 sound are likely to be inaccurate if based on a single assessment of perception, for the insights one gains may vary according to the protocol employed (see Flege & Hillenbrand, 1985a, for a discussion). For example, the results of a forced-choice identification task may lead
to a different picture of L2 perception ability than an AX same-different task (see Williams, 1979, 1980). Barry (1978) noted that, in an AX discrimination test, the percentage of pairs correctly discriminated by L2 learners varied as a function of inter-stimulus interval.

As noted above, careful consideration must be given to the context of the L2 sound being perceived. In examining identification by Japanese learners, Shimizu and Dantsuji (1983) found that English /r/ and /l/ were identified better in the context of front than back vowels (about 90% versus 80% correct), and better in isolated words than words embedded in a carrier phrase. The authors suggested this was due to a “more complete” articulatory specification of sounds in isolated words, but it is also possible that the demands of processing a longer stretch of speech makes it more difficult for listeners to identify new L2 sounds (see Flege, 1984a). Support for this hypothesis was provided by Winitz (1981), who found that discrimination of Chinese tones was more accurate in isolated syllables or disyllables than in short sentences.

INTERLINGUAL IDENTIFICATION

Investigators (e.g., Wode, 1976, 1977, 1978) have hypothesized that L1 sounds are used to produce L2 words because sounds in L1 and L2 are identified with one another. The term “identify” implies that a sound occurring in L2 is judged to belong to an L1 category, just as if the L2 sound were an allophone of some L1 phoneme. Weinreich (1953, p. 14) noted that when the learner identifies an L2 sound with a sound in L1, the L2 sound tends to be replaced by the corresponding L1 sound (cf. Monod, 1971, p. 88). Valdman (1976, p. 38) observed that:

The articulatory habits of the foreign language partially overlap those of the native language, and the student must learn to make new responses to stimuli which are interpreted as identical to native language stimuli. For instance, French and English /s/ differ with regard to place of articulation. The former is a dental, and the later is an alveolar. The partial similarities he perceives in the acoustic signal of French /s/ will lead an English speaker to respond with the alveolar rather than the dental sound.

Two studies have indicated that the labeling of L2 vowels is largely predictable on the basis of the acoustic specification of vowels in L1. Scholes (1968) had six non-native speakers of English classify a matrix of 69 synthetic vowels differing in terms of the frequency of F1 and F2. Subjects indicated which vowel they heard by writing a key word, first in their native language (Korean, Farsi, Greek, Russian) and then in English. For five subjects, labeling bore little resemblance to vowel categorization by native English listeners, but instead conformed closely to the responses the subjects had given when labeling the stimuli using L1 vowel categories.

Wiik (1965) examined the categorization of English vowels by Finnish children aged 11 to 12 years. The children used Finnish orthography to write English words. This probably provided a good indication of which Finnish vowels were identified with the vowels in English words since letters stand in a nearly one-to-one relationship with sounds (and presumably categories) in Finnish. The percentages of Finnish vowel responses given for English monophthongs are presented in Table 1.

Two findings are worth noting. First, the children were consistent in using Finnish categories for English vowels, as would be expected if interlingual identification occurs. Second, their responses were predictable in nearly every instance from an examination of the F1 and F2 values in English and Finnish vowels. For example, acoustic measurement of Finnish long and short /o/ indicated that these vowels occupy about the same portion of the acoustic phonetic vowel space as English /u/. As shown in Table 1, the Finns nearly always identified English /o/ as Finnish long or short /õ/, English /u/ has a higher average F2 value than its Finnish counterpart, which means that it occupies a portion of the acoustic vowel space occupied by Finnish /y/ and /õ/. As a result, the Finnish subjects identified Finnish /l/ and sometimes Finnish /y/ with English /l/. To take a final example, English /l/ occupies portions of the acoustic vowel space occupied by Finnish /õ/, /æ/, and /l/. Not surprisingly, the Finnish subjects identified English /l/ with all three Finnish vowel categories. Finnish /õ/ responses were most frequent, probably because this vowel is acoustically most similar to English /l/.

| Table 1. The Percentage of Times Finnish Vowel Categories Were Used to Represent the Vowels in English Words (after Wiik, 1965). |
|-----------------|-----------------|-----------------|-----------------|
|                | i   | e   | y   | õ   | æ   | a   | o   | u   | other |
| English target |     |     |     |     |     |     |     |     |       |
| vowel          |     |     |     |     |     |     |     |     |       |
| æ              | 90  | 2   | 85  | 9    | 3    |     |     |     |       |
| a              | 2   | 97  | 3   |     |     |     |     |     |       |
| o              | 17  | 80  | 3   |     |     |     |     |     |       |
| U              | 1   | 1   | 20  | 77   | 1    |     |     |     |       |
| u              | 9   |     | 86  | 4    |       |     |     |     |       |
| õ              | 98  |     |     |       |       |     |     |     |       |
| /õ/            |     | 2   | 59  | 16   | 10   | 2    |     |     |       |
Another study provided evidence that L2 learners identify L1 consonants with consonants in L2. Wode (1977) described the English pronunciation of his four native German children, who were learning English naturally. Wode focused on the production of /r/, a sound realized as a uvular fricative in German. The children used a German /r/ in English words before mastering English /r/ (an approximate realized in a variety of ways, usually involving use of the tongue tip or dorsum).

Four observations suggested that the German children identified German /r/ with the /r/ of English. It is unlikely they heard native English speakers using a uvular /r/ in English words. After seeming to master English /r/, one of the children again began substituting German for English /r/. This suggests the possibility that this child first judged English /r/ to be a "new" sound (i.e., did not identify it with German /r/), but later judged it to be a realization of German /r/ (i.e., what we have been calling a "similar" sound). The children spontaneously commented on the realization of /r/ in English words (e.g., "Lars tried to tell Inga that 'Trinity Center' should be produced /r/ (i.e., American /r/—JEF) not /R/`). Finally, when the children used English words in German sentences they produced a uvular /r/ even though they produced the same word with an English /r/ when speaking English.

Interlingual identification can be inferred from imitation experiments. For example, Yeni-Komshian, Carramaza, and Preston (1977) had Arab subjects imitate a continuum of stop consonants differing in VOT. Only some of the synthetic CV syllables used had VOT values comparable to those in Arabic stops. The subjects tended to produce stimuli identifiable as /da/ with the lead (i.e., prevoiced) VOT values characteristic of Arabic /d/, and to produce stimuli identifiable as /ta/ with the short-lag VOT values characteristic of Arabic /t/. Stimuli with the long-lag VOT values typical of English but not Arabic /t/ were realized with short-lag VOT values. This finding suggests that, prior to organizing the motor response needed for imitation, the subjects identified stimuli as belonging to an L1 category despite the substantial VOT (and other) differences between the synthetic stops and stops in their L1.

The primary basis for interlingual identification seems to be the degree of acoustic similarity between sounds found in L1 and L2. However, despite many seemingly straightforward examples of interlingual identification, L2 research has not yet established valid and reliable methods for determining which L1 sound(s)—if any—are identified with sounds occurring in L2. Acoustic similarity between L1 and L2 sounds is probably not the only factor involved. For example, evidence from speech production suggests that Americans regard the /y/ sound in French words as belonging to their English /u/ category (see Walz, 1979; Flege & Hillenbrand, 1984). However, speakers of some West African languages seem to interpret French /y/ sounds as belonging to their /i/ category (N. Spector, personal communication). And what was once French /y/ is now implemented as an /i/-quality vowel in Haitian Creole (C. Ferguson, personal communication).

This evidence suggests that perceived articulatory (or visual) similarity may play an important role in determining which sounds are identified with one another. It also suggests that similarity judgments might be influenced by phonological structure. Catford (1968, p. 164) concluded that the only basis for identifying L1 and L2 sounds is "substantial . . . rather than formal." This implies that sensory information derived during the perception and perhaps production of sounds, but not cognitive structures, play a role in determining which L1 and L2 phones are identified with one another. The possibility exists, however, that adult L2 learners note correspondences between the system of phonological contrasts in L1 and what they perceive to be the phonological system of L2 (see, e.g. Weinreich, 1953/1963). Perhaps L2 learners attempt to find "pattern congruity" in the phonemic categories of L2 by drawing analogies between L1 and L2 phonemes.

It is possible, too, that L2 learners attempt to maintain learned patterns of distinctive feature oppositions important in L1 but not L2. Michaels (1974) hypothesized that the relative importance of distinctive features in the learner’s phonological representation of sounds will influence which L1 sound is identified with, and perhaps substituted for, an L2 sound(s). He illustrated this hypothesis with anecdotal evidence. Both Russian and Japanese have a /t/ and an /s/. Despite this, Russians are said to substitute their /t/ for English /θ/, whereas Japanese learners are said to substitute their /s/ for English /θ/. Michaels speculated that the perception of "non-stridency" in English /θ/ leads Russians to substitute the closest non-strident sound of Russian (i.e., /t/), whereas Japanese learners’ perception of "continuancy" in English /θ/ leads them to substitute the closest continuous sound of Japanese (i.e., /s/). Note that /t/ and /θ/ are both classified as "nonstrident," and that both /s/ and /θ/ are classified as "continuant" sounds. The implication is that the relative importance of distinctive features (or degree of “distinctiveness”) may differ across languages, and that these differences affect how speakers of different languages will produce L2 sounds.

One general cognitive process likely to lead to the interlingual identification of sounds is equivalence classification, something which occurs when two objects are classified as categorically identical despite the existence of detectable physical differences between them. Equivalence classification can be illustrated as follows. Several times in recent months
a friend has mistakenly identified a silver Honda Civic parked near my home as being my Honda. The two Hondas were judged "equivalent" because the observer failed to note physical characteristics that distinguished the two cars (e.g., a dent on the front fender of one). This was a reasonable error. People naturally learn to attend primarily to "distinctive" features of cars (e.g., Honda, Civic, silver, 1982 model) rather than "nondistinctive" features (e.g., small dents). A person's perception of Hondas might change if his or her interaction with the environment demanded a change (say, if one were to become an insurance adjuster), or if some other factor precipitated a search for nondistinctive features (for example, frequent exposure to two similar Hondas parked side by side).

Equivalence classification appears to be operative in L1 acquisition. Infants are able to group physically different sounds on the basis of similar phonetic attributes (Hillenbrand, 1983), and they readily identify the physically different sounds produced by various talkers in many different phonetic contexts as being the "same" despite sometimes large acoustic differences (Kubaska & Aslin, 1983). What remains to be determined is whether equivalence classification influences the phonetic perception of adults and, if so, whether additional experience or changes in attentional strategies can modify the effect of equivalence classification as suggested by the analogy just given.

In early stages of L2 learning, individuals may perceive all of the sounds of L2 in terms of L1 sounds. If so, native English speakers should consistently identify even highly "exotic" sounds such as the clicks in Southern Bantu languages, with a specific English sound. However, we made a distinction earlier between new and similar L2 sounds. This represents the implicit hypothesis that L2 learners may eventually recognize that certain L2 sounds are not acoustically different realizations (or "allophones") of L1 categories. In early stages of learning, an L2 learner might regard a new L2 sound (e.g., /ʃ/ for a Dutch learner of English, or /y/ for an English learner of French) as a "distorted" version of an already familiar (L1) category. At later stages, however, the learner might determine that new sounds represent a category which is distinct from any previously established category. If so, similar L2 sounds (e.g., English /s/ for Dutch learners; French /u/ for English learners) might continue to be identified with L1 sounds.

It is uncertain at present whether interlingual identification ceases as L2 learners become familiar with L2 sounds. This could be tested directly by gathering systematically similarity judgments between pairs of L1 and L2 sounds at different stages of learning. If L2 learners eventually develop a separate category for new sounds, the perceived similarity between a new L2 sound should decrease—perhaps precipitously—with increasing L2 experience. The perceived similarity of similar L2 sounds, on the other hand, should remain constant with increased L2 experience.

PHONOLOGICAL FILTERING

A phenomenon familiar to language teachers is the seeming inability of some learners to distinguish in speech production L2 sounds not assigned to different categories in L1. For example, the single liquid of Japanese (usually regarded as an /r/ sound) shares acoustic properties with the /r/ and /l/ of English. Spanish /i/ acoustically resembles the /i/ and, to a somewhat lesser extent, /i/ of English. As a result, Japanese learners of English often appear to substitute /r/ for /l/ and /l/ for /r/ in English words; and native Spanish speakers substitute /i/ for /i/ in English (Hammond, 1982). It appears that in such instances, the L2 learner perceives sounds that are phonetically distinct for English-speaking listeners (i.e., /i/, l/, and /r, l/) as nondistinct.

L2 researchers have long assumed that such speech production errors are perceptually motivated, the result of a special form of equivalence classification known as "phonological filtering." Bloomfield (1933, p. 79) proposed that individuals learn to attend to just those features of sounds that are linguistically distinctive, ignoring acoustic differences which are redundant or "non-distinctive" for phonological contrasts in L1. Michaels (1974, p. 69) conjectured that the phonology of L1 acts like "a filter which retains only those features of a sound which are relevant to the identification of a phoneme." For Trubetzkoy (1939/1969, p. 51 ff) the learner's phonological filter acts like a "sieve," passing only information that is maximally useful for deriving the categorical identity of sounds in L1.

Others have argued that nondistinctive features of sounds may be both perceivable and important to L2 learning. Brière (1966, p. 74) noted that it is necessary to consider in detail the articulatory and acoustic properties of L1 and L2 sounds if one is to predict effectively which L1 sound will be substituted for an L2 sound (see also Walsh, 1974). Even Jakobson (Jakobson & Halle, 1956, p. 74), one of several linguists with whom the theory of distinctive features is associated, noted that although speakers learn to respond primarily to distinctive features, they do not learn to ignore nondistinctive features, which may also be coded into the speech signal and be under the active control of the talker.

Stampe (1979) observed that L1 acquisition would be impeded were phonological filtering to operate in L1 acquisition, for it would prevent children from accurately appraising the acoustic phonetic properties of
sounds in their native language. This represents the implicit hypothesis (see also Redmond, 1977) that during LI speech development segmental perception shifts from the direct perception of all features of a sound to a focus on distinctive features.

As the term has been used, phonological filtering seems to imply the operation of both “interlingual identification” and “equivalence classification.” For phonological filtering to occur, it seems logically necessary to assume that not only does the learner process enough sensory information to recognize the overall physical similarity between certain LI and L2 sounds, but also judges them to be categorically identical as the result of ignoring physical differences that might serve to distinguish them.

Viewed in this way, phonological filtering has something in common with categorical perception, which is said to limit listeners’ ability to discriminate between stimuli whose physical differences are auditorily detectable. Speech stimuli are said to be perceived “categorically” when a listener’s ability to discriminate between any pair of speech sounds is predictable on the basis of how each was previously identified. Categorical perception predicts, for example, that two stimuli differing by 30 ms of VOT will not be discriminated reliably if both are identified consistently as /t/, whereas another pair of stimuli differing by 30 ms of VOT will be discriminated reliably from one another if one is labelled /d/ and the other /t/.

A question of importance to L2 learning is whether phonological filtering makes it impossible for L2 learners to perceive the properties of L2 sounds that are nondistinctive in L1 but distinctive for monolingual native speakers of L2. The hypothesis implicit in the concept of interlingual identification is that L2 learners may be unable to discriminate certain pairs of physically different L1 and L2 sounds if they have been identified as belonging to the same category.

At least one study indicated that listeners have difficulty discriminating vowels identified as belonging to the same category in L1. Jacob, Alfonso, & Maxon (1980) found that native English speakers had no trouble discriminating pairs of vowels (e.g., /i/-/i/ or /u/-/u/) in a CVC context. The members of each pair they heard consisted of vowels that are phonetically distinct in English. Native Hebrew-speaking subjects, on the other hand, showed accurate discrimination for pairs in which both vowels (e.g., /i/-/u/) or neither vowel (e.g., /u/-/u/) occurred in their L1, but were far less accurate in discriminating pairs in which just one member was a Hebrew vowel (e.g., /i/-/i/ or /u/-/u/). This finding suggested that the Hebrew subjects categorized both vowels in such pairs as the “same” even though the native English speakers were able to discriminate them accurately. The authors concluded, like Trubetzkoy, that linguistic experience leads the mature speaker to process only those acoustic cues “which carry some degree of linguistic relevance” in L1. One practical consequence of this finding is that certain L2 learners may be unable to discern the phonetic, but not necessarily acoustic, difference between vowels in L2.

Other studies, however, have suggested that L2 learners do not irreversibly “filter out” acoustic differences that distinguish similar L1 and L2 sounds when those differences are not relevant to categorical identity in L1. Kent (1973) had native English speakers imitate synthetic /i/-/a/ and /i/-/o/ continua. Subjects appeared to be slightly more accurate in imitating stimuli that fell between clear tokens of /i/ and /a/ than between clear tokens of /i/ and /a/. Kent speculated that this was due to the fact that there are several categories intermediate to /i/ and /a/ but not /i/ and /a/ in English. The subjects’ ability to imitate vowel stimuli intermediate to stimuli representing “clear” tokens of English vowel categories suggested that they could discriminate between stimuli that would be identified as belonging to a single L1 category.

This inference is supported by the findings of Stevens, Liberman, & Studdert-Kennedy (1969), who examined the labeling and discrimination of synthetic vowels by English and Swedish-speaking subjects. Subjects in both groups labeled members of a continuum of unrounded front vowels (/i/-/i/-/a/ and /i/-/a/) with similar consistency, whereas the Swedes were somewhat more consistent than the English subjects in labeling the members of a rounded vowel continuum. (Swedish, but not English, distinguishes between several categories of nonback rounded vowels.) Despite this result, both the English and Swedish subjects showed the greatest accuracy in discriminating the same pairs of adjacent vowel stimuli. The authors suggested that the especially good discriminability of /i/ versus /a/ stimuli for all subjects derived from innate auditory processing abilities rather than from prior linguistic experience.

The frequent finding of categorical perception experiments is that stimuli identified as belonging to different categories are discriminated more accurately that those identified as belonging to the same category. In English, unlike Swedish, there is no distinction between an /a/ and an /a/ category. The English subjects were unfamiliar with a phonetic contrast between pairs of vowels such as /i/-/a/ and /a/-/o/ because these pairs are not phonetically distinct in English. This did not prevent them, however, from perceiving the nondistinctive acoustic differences between those pairs of vowels.

Other counterevidence to the belief that phonological filtering makes it impossible for L2 learners to perceive accurately the sounds of L2...
relates to the perception of consonants. Yeni-Komshian, Preston, & Kewley-Port (1968) had subjects (American children and native Arab-speaking adults and children) imitate the members of a /da/-to-/ta/ continuum in which the VOT of the initial stop varied from -150 to +150 ms. Subjects tended to produce stimuli with two modal VOT values typical of stops in their native language, but there were some important differences. Some of the Arab children imitated stimuli with three distinct VOT values (i.e., lead, short-lag, and long-lag) rather than the two ranges of VOT (lead and short-lag) typical of Arabic. This suggested that they were able to discriminate sounds that would not be assigned different category labels in L1 since their L1 does not contrast short- and long-lag stops.

Some of the Arab adults seem not to have heard stimuli with VOT values of 50-150 ms as stops because they differed so noticeably from Arabic /!, which is realized with a VOT value that is significantly shorter than that of English (Flege, 1979). In imitating these stimuli, they often produced an Arabic fricative (e.g., /x/) in syllable-initial position. This suggested that they detected the acoustic difference between long-lag and short-lag stops even though this kind of VOT difference is not used in Arabic to distinguish categories of stops.

Flege and Hammond (1982) provided further evidence that L2 learners may not necessarily “filter out” properties of L2 sounds that are nondistinctive in L1. Their results suggested that native English speakers are able to distinguish short-lag tokens of /t/ produced in English words by native speakers of Spanish from tokens of /t/ produced with long-lag VOT values in English words by native speakers of English. English-speaking subjects who were familiar with Spanish-accented English were asked to read several sentences containing a word-initial /t/ with a “typical Spanish accent.” Subjects imitating a Spanish accent produced /t/ with VOT values about 40 ms shorter than native English subjects who read the same sentences without having been instructed to imitate a Spanish accent. The conclusion that the subjects were tacitly aware of purely phonetic (i.e., subcategorical) acoustic differences distinguishing the /t/ of standard and Spanish-accented English was supported by the finding that subjects who produced the most sound substitutions actually heard in Spanish-accented English (e.g., /bes/ for “vase” and /big/ for “big”) also shortened VOT the most.

Finally, Flege (1984a) found that listeners were able to distinguish the /t/ produced in English words by native French and native English speakers, even though all the sounds presented to listeners in identification and forced-choice comparison tasks were identifiable as /t/. This finding is consistent with the observation that native speakers are able to detect “distortions” of sounds (e.g., the /s/ produced by lispers), as well as to detect subtle phonetic differences between dialects of their native language.

Taken together, these studies indicate that listeners process a great deal more acoustic information present in the speech signal than is needed for categorization. This conclusion is consistent with the finding (e.g., Liberman, Harris, Hoffman, & Griffith, 1957) that within-category discrimination is better than predicted from category labeling, and the finding that it can be improved with training (e.g., Carney et al., 1977). Thus, on balance, we must conclude that phonological filtering does not make it impossible for L2 learners to perceive nondistinctive features of L2 sounds accurately.

**REINTERPRETATION**

It is possible that the degree of physical similarity between L1 and L2 sounds is more important to interlingual identification than the overall system of phonemic oppositions in L1 and L2 (see, e.g., Catford, 1968; F. Johansson, 1973, p. 10 ff). Yet it is also conceivable that the phonological significance of various properties of sounds plays some role in determining how L2 sounds are perceived and produced. Children acquiring L1 sometimes produce acoustic distinctions between categories that are strikingly dissimilar from those produced by their parents, either because they are unable articulatorily to implement the dimensions used by their elders, or because they don’t yet know which dimensions are linguistically important in L1 (Camarata & Gandour, 1985). Weinreich (1953, p. 18) believed that L2 learners sometimes also fail to perceive the linguistic significance of acoustic dimensions in L2 sounds. He hypothesized that dimensions which have perceptual import in L2 but not L1 may not be honored by L2 learners. The term Weinreich gave to the perceptual processing of L2 sounds in terms of L1 categories is “reinterpretation.”

Three speech production studies (Suomi, 1976; Elsendoorn, 1980; Mitleb, 1981) have lent indirect support to the hypothesis that L2 learners reinterpret the sounds of L2. In these studies, the temporal difference between tense and lax English vowels spoken by non-native speakers of English was measured. In English, tense vowels (e.g., /i/) are longer than their lax cogeners (/a/). For native speakers of English, the temporal difference between tense/lax pairs is generally much less important to categorization than the spectral differences which also distinguish tense from lax vowels (Bennett, 1968; Ainsworth, 1972). The native languages of the subjects studied (Finnish, Dutch, and Arabic) do not possess a distinction between tense and lax vowels, but
all three languages possess a distinction between phonemically long and short vowels (i.e., V: versus /V/).

Perhaps as a result, the non-native speakers produced an even larger temporal distinction between tense and lax vowels than native English speakers. The non-native speakers produced tense/lax English vowel pairs with temporal differences appropriate for the distinction between long and short vowel phonemes in their L1. One explanation for the non-native speakers’ seeming exaggeration of the temporal contrast between tense and lax English vowels is that they reinterpreted the English tense/lax distinction as a phonemic length distinction.

The spectral differences between phonemically long and short vowels in quantity languages such as Finnish, Dutch, and Arabic appear to be much less prominent acoustically, and less important perceptually, than differences in duration (Lehiste, 1970). As the result of prior L1 experience, the Finnish, Dutch, and Arabic learners of English may have regarded the temporal difference between tense and lax vowels in English as more important than the spectral differences between them. If re-interpretation did occur it would mean that the apparent success of the Finnish, Dutch, and Arabic learners in producing a new temporal distinction in L2 was not actually the result of learning a new temporal distinction (i.e. the English tense/lax distinction) or modifying a pre-established (L1) pattern of speech timing. Further research is needed in which vowel quality and duration are orthogonally varied to test directly the hypothesis that reinterpretation occurred.

PERCEPTUAL TUNING

Neither experience with specific speech sounds nor the development of innate mechanisms would be necessary for the child to perceive the sounds of L1 accurately and efficiently if speech perception depended on innate or highly specialized auditory "feature detectors." However, this view has been questioned repeatedly in recent years. Researchers recognize increasingly that the auditory processing of speech events are shaped, within limits imposed by the auditory system, by experience with specific varieties of speech (see, e.g., Strange & Jenkins, 1978; Summerfield, 1982).

Aslin and Pisoni (1980) speculated that a "tuning" mechanism underlies the human ability to perceive optimally the phonetic attributes of L1 sounds. The existence of such a mechanism is consistent with the classical view that, in perceptual development, the organism follows a path which gradually increases its ability to extract and use environmentally important information (Gibson, 1969). Moreover, tuning would enable species that depend on vocal signals, such as humans and songbirds, to reduce the amount of genetic pre-programming necessary to ensure adequate and authentic vocal behavior. At the same time, it would make possible the development of "dialects" based on experience with specific environmental stimuli (Marler & Mundinger, 1971; Eimas & Tartter, 1979; Gottlieb, 1976).

The question of interest in the present context is whether the perceptual system, once optimally tuned for the processing of sounds occurring in L1, can be "retuned" again during the process of L2 learning for the optimal processing of new and similar L2 sounds. Research in behavioral embryology and visual perception suggests a number of possible effects of experience on the development of perceptual systems. According to Aslin and Pisoni (1980), the ability to perceive important environmental stimuli may follow three possible courses. First, a perceptual ability might be undeveloped at birth, in which case it would either be induced by appropriate stimulation, or remain undeveloped in the absence of appropriate stimulation. Second, a perceptual ability might be present at birth in a fully developed form, in which case it would either be maintained through experience or lost in the absence of relevant stimulation. Third, a perceptual ability might be partially developed at birth. If so, relevant post-natal experience (or lack thereof) would serve to maintain the perceptual ability in its original form, facilitate further development, or induce its loss.

The results of infant speech perception research largely rule out the first possibility. Extensive evidence of cross-language differences in speech production suggests that the second possibility is untenable. The third possibility outlined above sets the most plausible agenda for developing and testing hypotheses concerning perceptual development in L1 speech acquisition. As we will see, the literature also provides evidence for the maintenance, learning, and possible loss of perceptual abilities during L2 learning.

**Maintenance.**

Maintenance seems to characterize the perception of L2 sounds in early stages of L2 learning. Individuals tend to label L2 sounds using categories established for L1 sounds. This can be observed by presenting the same sound to listeners differing in L1 background. If the sound is realized differently in the listeners' native languages, it may evoke different perceptual responses. For example, a labial stop realized with short-lag VOT values (i.e., a voiceless unaspirated stop) will usually be labeled as /p/ by native speakers of Spanish, but as /b/ by native speakers of English (Elman, Diehl, & Buchwald, 1977; Williams, 1977a).
Loss.

Studies of perceptual development often reveal that an organism learns to attend just to stimulus information that is relevant to categorization, and to ignore information that is either redundant or not relevant to categorization (Gibson, 1969). There is evidence that humans may lose the ability to identify and discriminate speech sounds not found in their L1, or become less sensitive to L1 sounds as the result of learning L2 sounds.

Previous research has revealed instances in which listeners fail to discriminate or identify L2 sounds as accurately as native speakers of L2. In one study (Marckwardt, 1944; see also Marckwardt, 1946), Spanish L2 learners were reported to identify the /tʃ/ and /ʃ/ sounds in English words like “church” and “shoe” incorrectly about 20% of the time. Their difficulty seems to have had both a phonological and phonetic motivation. There is no /ʃ/ in Spanish, and English /tʃ/ is produced with a longer period of fricative release than its Spanish counterpart, so that it might seem to Spanish speakers to resemble English /ʃ/.

Another study (Oakeshott-Taylor, 1976) examined the ability of German learners to identify sounds that are contrasted in English but not German, such as voiced and voiceless word-final obstruents, and the vowels /ɛ/ and /e/. Subjects identified incorrectly the voice feature in obstruents about 17% of the time and their vowel identification errors averaged 27%. The subjects seemed to favor use of the response alternative representing a category not found in their L1 (e.g., word-final /b,d,g/ rather than /p,t,k/; or /ɛ/ rather than /e/). This apparent bias may have resulted from their explicit knowledge that native speakers of German have difficulty producing English sounds such as /b/, /d/, /g/, and /ɛ/. That is, they may have been overcompensating for a known difficulty.

Perhaps the best known loss of a specific speech perception ability pertains to the perception of English /r/ and /l/ by native speakers of Japanese. Japanese learners often fail to produce a perceptually effective contrast between /r/ and /l/ in English words because there is no phonemic contrast between these sounds in Japanese. Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, (1975; but cf. Shimizu & Dantsuji, 1983) found that adult Japanese subjects were unable to discriminate between pairs of stimuli differing in an acoustic dimension (third formant onset frequency) sufficient to cue the /r-l/ contrast for native speakers of English. Native English subjects tended to discriminate accurately only those stimuli they labelled differently. This is the pattern predicted by “categorical” perception.

The Japanese subjects, on the other hand, discriminated all pairs at only slightly better than chance levels, including those given different category labels by native speakers of English. The lack of enhanced discriminability at the /r-l/ phoneme boundary suggested that the Japanese listeners did not assign different category labels to members of the continuum. This interpretation was supported by the results of a control experiment using non-speech stimuli. Japanese and English subjects did not differ in discriminating isolated third formants which, when presented along with the F1 and F2, were sufficient to cue the contrast between /r/ and /l/. Moreover, pre-articulate infants are able to discriminate /r/ and /l/ stimuli, presumably on an auditory rather than phonetic basis (Eimas, 1975).

Analogous results were reported by Werker, Gilbert, Humphry, & Tees (1981) concerning the ability of native English-speaking adults to discriminate pairs of naturally produced Hindi sounds. Unlike adult native speakers of Hindi, native English adults were unable to discriminate dental versus retroflex, and breathy voiced versus voiceless aspirate dental stops which occur in Hindi. Six-month-old infants being raised in English-speaking homes, on the other hand, were able to discriminate the Hindi sounds.

This finding is consistent with earlier studies (e.g., Eimas, 1975) that have demonstrated infants’ ability to discriminate auditorily between many, if not all, of the speech sounds categories occurring in human languages. It suggested that the infants, unlike the English adults, had not “learned” to ignore the acoustic properties of Hindi stops that Hindi speakers use to discriminate sounds in their L1. Findings such as these seem to show that perception is related to previous experience. However, they do not make it clear whether an ability has been lost or whether a specific ability has simply not yet been developed.

One line of research suggests that perceptual differences between groups of individuals differing in language background follow from the development or enhancement of perceptual abilities. Eilers, Gavin, and Oller (1982) found that babies aged 6 to 8 months who were being reared in English-speaking homes were significantly poorer in discriminating a Spanish contrast (a tap versus trilled /r/) than babies being reared in Spanish-speaking homes. The authors concluded that early linguistic experience enhances the ability to distinguish at least some of the sound contrasts found in human languages.

In a later study, Oller and Eilers (1983) found that two-year-old children were better able to discriminate between two LI (Spanish or English) sounds than a pair of sounds that included one LI sound and one sound not found in LI. Spanish-learning children had relatively great difficulty distinguishing /w/ from the approximate /r/ which occurs...
in English but not Spanish, and English-learning children had relatively great difficulty distinguishing a tap from a trilled /r/ (the second of which does not occur in English).

Several other studies support the belief that specific speech perception abilities are lost in the absence of environmental stimulation. In a longitudinal study, Werker and Tees (1983) tested the ability of infants to discriminate the contrast between unaspirated dental versus retroflex Hindi stops, and the velar versus uvular glottalized stops found in Thompson (a Salish language). The infants discriminated both non-English sound contrasts at the age of six months, but were no longer able to do so at twelve months of age. The authors concluded that speech perception abilities may diminish at an early age in the absence of specific linguistic experience (see also Trehub, 1976).

Bond and Adamescu (1979) trained adults and children 4, 6, and 12 years of age (6 per group) to identify objects that had been given names beginning with a plosive stop (like those found in English) or an implosive stop (such as those found in Hausa). The four-year-olds, but not the twelve-year-olds or adults, performed at above-chance levels in distinguishing this non-English phonetic contrast. This suggested that only the youngest subjects had retained the sensitivity needed to discriminate sounds distinguished by a phonetic contrast not exploited by their L1.

Findings reported by Shimizu and Dantsuji (1983) suggested that the loss of ability to discriminate sounds not found in L1 may occur after the age of 5 years. They found that Japanese adults, and to an even greater extent Japanese five-year-olds, showed a pronounced peak in correct discrimination near the /r/-/l/ category boundary (established by means of a forced-choice identification experiment for native English speakers). Since the Japanese children were tested in Japan and did not speak English, it is unlikely that their discrimination peak resulted from exposure to the English /r,l/ contrast. It is possible that, for the Japanese children, evidence of enhanced discriminability for certain pairs of stimuli in the /r-l/ continuum was the result of some nonlinearity in peripheral auditory processing rather than phonetic processing, just as it presumably is for infants (e.g., Eimas, 1975). Although it is likely that by the age of five years children process speech in a primarily phonetic rather than auditory mode, it is conceivable that children are better able than adults to switch to an auditory mode of processing, and that this ability diminishes with age.

Several studies have indicated that learning L2 may reduce sensitivity to phonetic contrasts in L1. Garnes (1977) examined the labeling of stimuli differing in both vowel and fricative duration by native speakers of Icelandic. The Icelandic word is (ice) contains a phonemically long vowel (/V:/) that is about twice as long as the phonemically short (/V/) vowel in iss (of ice). Icelanders who had never lived outside Iceland, as well as nine Icelanders who had resided for nine months or more in the United States, showed a clear crossover from iss to is judgments as vowel duration was increased systematically. However, there were significantly more "ambiguous" stimuli (i.e., stimuli not labeled as iss or is more than 75% of the time) for subjects who had lived in the United States than for those who had not. One possible interpretation of this finding is that massive exposure to English modified the Icelanders' use of vowel duration as a cue to the phonemic length distinction because, in English, vowel duration is conditioned by many factors but is not used to contrast long versus short vowel phonemes.

This interpretation is supported by a finding reported by Gottfried and Beddor (1984). It appears that native speakers of English are influenced by temporal variations to a greater extent in identifying /o/ and /a/ than native speakers of French, who rely almost exclusively on spectral quality. Gottfried and Beddor (1984) found that experienced but not relatively inexperienced English speakers of French resembled French native speakers in showing little effect of duration in identifying the members of an /o/-to-/a/ continuum. That is, those native English speakers who were most experienced in French did not seem to use vowel duration as a secondary cue to a phonetic distinction between vowels.

Williams (1977a, 1977b, 1979, 1980) examined the perception of English /p/ and /b/ by native Spanish speakers who had learned English. Each of several groups of Puerto Rican children showed enhanced discrimination of stimuli straddling the English phoneme boundary. The magnitude of the peak increased as a function of length of residence in the United States. Enhanced discriminability of stimuli straddling the Spanish phoneme boundary between /p/ and /b/ on the other hand, seemed to diminish with age. Unlike Spanish monolingual adults, only two of eight adult Spanish speakers of English showed a discrimination peak at the Spanish phoneme boundary. Subjects aged 14 to 16 years showed a much less prominent discrimination peak at the Spanish than English phoneme boundary; and subjects aged 8 to 10 years showed no evidence of a discrimination peak at the Spanish phoneme boundary.

One possible explanation for this finding is that the English acoustic distinction between short-lag and long-lag stops is easier to discriminate auditorily than the Spanish acoustic distinction between prevoiced and short-lag stops. As noted by Williams (1977b, pp. 101-102), the English contrast between voiced and voiceless stops in prevocalic position is
richer in acoustic cues than the corresponding Spanish stop contrast. Moreover, Macken (1980) noted that children acquiring Spanish seem to establish a VOT distinction between voiced and voiceless later than children acquiring English. This seems to happen because categories of stops are reliably distinguished by VOT in prevocalic position, but a VOT distinction occurs only variably and in restricted phonetic contexts in Spanish. Perhaps L2 learners, especially young ones, readily abandon the tendency to focus attention on acoustic dimension(s) distinguishing L1 sounds if the acoustic dimension(s) distinguishing similar L2 sounds is easier to process, or more reliably present.

Learning.

Many studies have indicated that perceptual learning occurs during L1 acquisition. For example, S. Johansson (1975) found that Swedes develop a sensitivity to the properties of final stops which native English speakers appear to lack. He manipulated the duration of the final stop closure interval in naturally produced Swedish and English words. Rating scale judgments of authenticity indicated that there was a narrow range of acceptable closure durations for the Swedish but not English subjects (see also Jonasson & McAllister, 1972).

Johansson (1975) also found that Swedish subjects' rating scale estimates of authenticity decreased sharply when an English burst was appended to the end of a Swedish word. Native English-speaking subjects, on the other hand, showed only a slight downward shift in ratings when a Swedish burst was appended to an English word. This suggested that the English subjects were less sensitive to the acoustic properties of stop release bursts than the Swedish subjects because English stops are released less regularly than Swedish stops. A similar conclusion was reached by Flege and Hillenbrand (1987) concerning the importance of the release burst of word-final stops for native speakers of French and English.

Flege and Hillenbrand (1983) examined the labeling of synthetic CVC stimuli differing in vowel and final fricative duration. Increases in vowel duration and (to a lesser extent) decreases in final fricative duration significantly increased the percentage of /z/ (i.e., “peas”) judgments by English and French subjects. The manipulation of vowel, but not fricative, duration affected the final voicing judgments of Swedish and Finnish subjects, even those who had lived for a year or more in an English-speaking environment. Final /s/ and /z/ in English and French are distinguished by both noise duration and the duration of a preceding vowel. Neither Swedish nor Finnish possess a contrast between final /s/ and /z/. These results therefore suggested that the French and English subjects had developed the ability to integrate temporal information pertaining to vowel and fricative duration on the basis of exposure to words in their L1.

It appears that perceptual learning also occurs during L2 learning. Schulman (1983) had Swedish speakers of English label the members of a vowel continuum using Swedish and English vowel categories. The subjects, most of whom had lived in an English-speaking country, demonstrated tacit awareness of some of the differences which distinguish vowels in their two languages. For example, they labelled more stimuli as /i/ when using English than Swedish categories.

Several studies have examined the identification and discrimination of stops differing in VOT. These studies have indicated that the perception of prevocalic stop consonants differing in VOT is not irreversibly constrained by previous experience, but evolves during L2 learning. Caramazza, Yeni-Komshian, Zurif, & Carbone (1973) reported that the phoneme boundaries between voiced and voiceless stops occurred at shorter values for French learners of English than those observed for monolingual speakers of English (29 ms), but at longer VOT values than those observed for monolingual speakers of French (14 ms). Similar results have been reported by Albert and Obler (1978; Obler, 1983) for Hebrew speakers of English who learned both Hebrew and English before the age of six and judged themselves to be equally competent in both languages.

Williams (1980) examined the perception of adult bilinguals who learned their L2 (either Spanish or English) early in childhood and pronounced both of their languages with a high degree of authenticity. Five of the eight subjects showed a cross-over from predominantly voiced to voiceless judgments at about the same VOT values (24 ms) as monolingual English subjects. This represented a large shift away from the boundary established for monolingual Spanish speakers (8 ms of VOT), in the direction of boundaries typical for native English speakers. All of the subjects showed enhanced discrimination at the English phoneme boundary. Several revealed what might be considered a “bilingual” pattern of perception, that is, enhanced discrimination of stimuli straddling both the Spanish and English phoneme boundaries.

Williams (1980) also examined the perception of Spanish children (8 to 10 and 14 to 16 years of age) who were learning English as a foreign language. The children’s phoneme boundaries occurred at shorter VOT values than monolingual English speakers, but at longer values than those of monolingual Spanish speakers. Some of the older children, like some of the adult Spanish/English bilinguals mentioned earlier, showed evidence of enhanced discrimination at both the Spanish and English phoneme boundaries. The younger children, on the other hand,
generally showed enhanced discrimination in the region of the English phoneme boundary between /p/ and /b/.

Several studies have examined the ability of individuals to learn a phonetic contrast not present in their L1. Barry (1978) examined the ability of L2 learners to distinguish the prevoiced (i.e., lead) /da/ from short-lag /ta/ found in French words. Identification errors were surprisingly high (about 32%). My analysis of data reported for 34 subjects randomly chosen from Barry’s Tables I and II indicated that students who had taken a French pronunciation course discriminated correctly a significantly larger number of stimulus pairs (77%) than those who had not (72%, p < .05). Interestingly, discrimination was significantly less accurate when there was a relatively long (2.5 sec) compared to short (0.5 sec) inter-stimulus interval. The decrement in correct discrimination brought about by increasing the ISI was somewhat greater for experienced (6%) than inexperienced subjects (1%), suggesting that the inexperienced subjects relied on auditory as opposed to phonetic information to a greater extent than the experienced subjects.

Other studies have examined the effect of short-term laboratory training on the perception of L2 sounds. For monolingual speakers of Russian, stimuli with VOT values of +10 ms as well as those with values of +60 ms would be heard as /p/ since Russian /p/ is produced with short-lag VOT values of around +10 msec. Lisker (1970) reported that discrimination training enabled native speakers of Russians to label stimuli with +60 ms VOT values as /pa/ and stimuli with +10 ms VOT as /ba/ (albeit not as consistently as native English speakers).

Another study (Pisoni, Aslin, Peres, & Hennessy, 1982; but cf. Strange & Jenkins, 1978) provided evidence that native English-speaking subjects can be trained to differentiate stops with lead and short-lag VOT values, even though this acoustic difference is not used in English to distinguish voiced from voiceless stop categories. Training consisted of the alternating presentation of “good” exemplars of “mba,” “ba” and “pa” categories (i.e., stimuli with –70, 0, and +70 ms VOT, respectively). Another study (McCasky et al., 1983) replicated this finding, and also demonstrated a generalization of the training effect to an untrained place of articulation.

Several studies have examined the perception of English /r/ and /l/ by native speakers of Japanese. Mochizuki (1981) examined the perception of Japanese subjects who had greater exposure to English than the subjects studied earlier by Miyawaki et al. (1975). These relatively experienced subjects were able to label natural and synthetic /r/ and /l/ stimuli accurately, albeit with somewhat less consistency than native English speakers. MacKain, Best, & Strange (1981) found that Japanese L2 learners may learn to perceive the English /r/-/l/ distinction in a native-like fashion as the result of massive exposure to English. Japanese learners with relatively great experience speaking and hearing English perceived stimuli in a continuum ranging from /r/ to /l/ in the same “categorical” fashion as native English-speaking subjects. That is, they showed a sharp crossover from /r/ to /l/ in the same “categorical” fashion as native English-speaking subjects. This crossover point was reached during the first 15 training sessions.

Two studies examined the effect of short-term training on the perception of /r/ and /l/. Strange and Dittman (1981) attempted to train eight Japanese subjects to discriminate English /r/ and /l/ using discrimination training with immediate feedback. Discrimination improved slowly but steadily during 15 training sessions and, as a result, post-training identification was better. However, there was little evidence of transfer of training to another synthetic speech continuum or to naturally produced /r/ versus /l/ minimal pairs. Cochrane (1977) administered five hours of /r/-/l/ training to adult and child native speakers of Japanese. The training consisted of instructions concerning tongue placement, as well as imitation by subjects of the /r/ and /l/ produced by an English native speaker in progressively longer stretches of speech. The pronunciation of adult but not child subjects was observed to improve. The lack of a training effect for the children might simply have reflected a ceiling effect.

Several other studies have indicated that the perception of temporal cues for the contrast between voiced and voiceless obstruents may change as the result of L2 learning. Elsendoorn (1981) investigated Dutch subjects’ use of vowel duration as a perceptual cue to the voicing contrast between word-final voiced and voiceless stops. The stop voicing judgments of monolingual speakers of Dutch is not influenced by preceding vowel duration, as it is for native English speakers. Elsendoorn found that Dutch L2 learners with relatively great experience in English (university students) but not less experienced Dutch subjects (high school students) showed an effect of vowel duration resembling that of native English speakers.

A technique employed in another study (Elsendoorn, 1984) permitted subjects to signal their perception of the ideal duration of vowels in English words without actually speaking. Subjects were asked to adjust the duration of English CVC words until they sounded “correct.” They represented vowels as being longer before /b,d,g/ than /p,t,k/. This suggested to Elsendoorn they had developed an “inner” knowledge that vowels should be longer before /b,d,g/ than /p,t,k/ in English words.

FOREIGN LANGUAGE SPEECH SOUNDS 297
Flege (1984b) examined the labeling by two groups of Saudi Arabians of a “piece” to “peas” continuum in which vowel and fricative duration were varied orthogonally. It is unlikely that monolingual speakers of Arabic will exploit either vowel duration or final fricative duration as a perceptual cue to the /s/ versus /z/ contrast since neither temporal dimension is used to distinguish /s/ from /z/ in Arabic words. Both groups of Arabs labeled stimuli with relatively long vowels as “peas” (i.e., final /z/), but only the most experienced subjects (2 to 4 years of residence in the United States) resembled native English speakers in labeling stimuli with relatively long fricative durations as “piece” (i.e., final /s/).

PHONETIC CATEGORY PROTOTYPES

It has been proposed that an important aspect of segmental phonetic perception is the comparison of the auditory properties of a sound being processed to an array of sensory properties stored centrally in phonetic category prototypes (Oden & Massaro, 1978; Samuel, 1982). These phonetic representations are thought to contain distinctive and nondistinctive information pertaining to the temporal and spectral characteristics of a phonetic category, as well as probabilistic information pertaining to the likelihood that a property will occur, or co-occur with some other property. Nooteboom (1973) conceptualized phonetic prototypes as “mental images” representing all language-specific aspects of “inner speech”. However, it is not possible at present to provide a detailed description of the kind of information stored in phonetic category prototypes, nor how that information might be structured.

Based on a wide variety of behavioral evidence, it seems likely that prototypes include a modal information related to the auditory properties of sounds, as well as information that may be translated into motor control parameters used for the movement and placement of the speech articulators (MacNeilage, 1970, 1980). The assumption made here is that prototypes contain at least enough information pertaining to timing, coarticulation, and place and manner of articulation to permit talkers to realize L1 sounds authentically. We also assume that prototypes contain enough information for listeners to detect auditorily and perceive cross-dialect and cross-linguistic differences in segmental articulation. One further assumption is that prototypes contain information pertaining to both sensory and motor characteristics of categories, and means for relating the two.

If speech perception is mediated by phonetic category prototypes, it is likely that learners will perceive an L2 sound differently from L1 native speakers when sounds are realized differently in L1 and L2. Flege (1984a) hypothesized that the changes in perception observed during the process of L2 learning occur as the result of changes in phonetic category prototypes. In the present context, we ask “How do prototypes develop?”

It is possible that infants are endowed with certain general prototypes containing information that is neither learned post-natally nor language-specific. Infants can discriminate many of the sound categories found in human languages prior to extensive exposure to, or use of, speech. In the beginning stages of L1 speech acquisition, sounds are probably categorized largely on the basis of a small number of invariant properties which are determined by non-linearities in auditory processing (Stevens, 1983; Kuhl, 1980; Kewley-Port, Pisoni, & Studdert-Kennedy, 1983). Supporting this contention is the recent finding that six-month-old infants are more sensitive to the properties of vowels subjectively judged by adults to be “good” (i.e., prototypical) than “poor” exemplars of a vowel category (Grieser & Kuhl, 1983). The ability to relate motor gestures to sensory information associated with a speech gesture also appears to be present at birth (Spelke, 1976; Meltzoff & Moore, 1977; Kuhl & Meltzoff, 1982).

The progressive elaboration of innate prototypes probably depends on massive experience speaking and hearing the sounds of a particular language. During L1 acquisition, the learner encounters an enormous quantity and variety of realizations of each speech sound category. The child may elaborate existing prototypes and establish new ones in order to ensure perceptual constancy in the face of the great acoustic variability evident for any category due to differences resulting from identity of the talker, phonetic context, speech rate, and so on. Information stored in prototypes might also help ensure that variations which do not follow intrinsically from properties of the speech production mechanism are encoded in a form appropriate for motor control.

Work in visual pattern recognition (e.g., Posner & Keele, 1968, 1970) has suggested that prototypes are formed automatically when viewers are exposed to many different exemplars of a visual pattern or form, and that experience with a wide variety of patterns may be useful for establishing and defining these prototypes. Language learners also seem to readily abstract phonetic category prototypes from their experience with specific realizations of a category. In Russian (as in other languages) the quality of vowels may differ as a function of phonetic context. A palatal consonant tends to raise the F2 of a preceding vowel. Crothers and Suppes (1967) found that English-speaking subjects learned to discriminate Russian vowels faster when presented with vowels occur-
ring in both palatalized and nonpalatalized contexts than when trained on vowels found in only a single context.

Investigating children’s categorization of three-dimensional objects, Mulford (1979) found that older children correctly categorized more objects than younger children, and that younger children showed a tendency to categorize correctly only those objects with a relatively large number of “core” features (that is, those with the largest number of features that were always and only found in the abstract “prototype” thought to underly the category).

Phonetic category prototypes may include progressively more information pertaining to the language-specific features of a sound as they are elaborated. If so, we would expect accent detection to improve as the prototypes are elaborated. This prediction was supported by results reported by Scovel (1981). Scovel presented 8-sec excerpts of speech produced by native and non-native English talkers to several groups of listeners. He noted higher rates of accent detection by older compared to younger child listeners who were native speakers of English. Moreover, he found that native English-speaking listeners were better able than non-English-speaking listeners to detect the presence of accent; and that advanced learners of English were better at detecting accent in English speech than beginning learners of L2.

Prototypes are thought to represent systematically sound properties that may vary according to factors such as talker, speaking rate, and phonetic context. If so, listeners should become progressively more adept at correctly identifying sounds as they gain experience with a sound as it is realized in a particular variety of speech. This hypothesis has received support from several studies. Elliott (1979) found that older children were better able than younger children to recognize L1 words in a predictable semantic context that had been degraded by multi-talker babbling. Parnell and Amerman (1979) found that older children and adults were better able than young children to recognize L1 syllables that had been altered by an electronic gating technique. Neuman and Hochberg (1983) found no difference between adults and children for speech produced in nonreverberant conditions, but did observe superior performance for adults than children for speech produced in reverberant conditions. McGarr (1981) found that listeners who were familiar with the English spoken by deaf individuals comprehended more of it than listeners unacquainted with deaf speech patterns.

L2 studies point in the same direction. It does not appear, as once thought, that the perception in noise of L2 provides a good indication of overall L2 competence (Johansson, 1975). However, there is clear evidence that noise does interfere more with L2 perception than L1

perception. A cross-sectional study by van Balen (1980) examined the comprehension by Dutch high school students of excerpts of L1 (Dutch) and L2 (English) speech. Intelligibility was always greater for Dutch than English excerpts, both when the speech was degraded by noise and when it was not. Intelligibility was found to increase systematically with English-language experience. There seemed to be a weak relationship between the length of the excerpt in Dutch but not English. This suggested that listeners were better able to exploit expectancies generated by syntactic and semantic regularities in L1 than L2.

Oyama (1982b) compared the ability of native speakers of English and Italian to comprehend short English sentences presented in noise. Comprehension was correlated negatively with age of first exposure to English. Comprehension by the non-native speakers was inferior to that of native English-speaking subjects for all except those subjects who arrived earliest and stayed longest in the United States. Johansson (1978) also reported lower intelligibility scores for non-native speakers (Swedes) than native English speakers. The effect of noise and quantizing error was greater for non-native than native speakers (see also Lehtonen & Leppanen, 1980).

At present there is little empirical evidence bearing directly on the time-course or nature of prototype development, either in L1 acquisition or L2 speech learning. Schouten (1975, p. 16) speculated that the elaboration of prototypes depends on the number of times a sound is produced and heard. He also claimed that the better defined a prototype, the more “every instance of perception and production will tend to gravitate towards it.” Based on these assumptions, he hypothesized that exposure to L2 sounds judged to be realizations of a category defined by a prototype established during L1 acquisition will have progressively less effect on it. Put another way, the extent to which a prototype changes during L2 learning will be inversely proportional to previous phonetic experience, so that prototypes for child learners of L2 should be more malleable than those of adult L2 learners.

An important goal for L2 research is to determine to what extent the phonetic category prototypes of native and non-native speakers differ. There is a need to determine under what conditions a prototype evolves as a function of linguistic experience, to what extent production of L2 sounds influences the elaboration of prototypes, and whether separate prototypes can exist for similar sounds in L1 and L2. Based on the studies reviewed earlier, four hypotheses regarding the effect of L2 learning on the development of phonetic category prototypes can be formalized for pairs of similar L1 and L2 sounds.

First, there might be no modification of existing phonetic category prototypes as the result of exposure to similar L2 sounds. If so, learners
might simply perceive L2 sounds as if they were L1 sounds. Such a pattern might exist in the earliest stages of L2 learning. However, the evidence reviewed earlier (see Elman et al., 1977) suggested that this will not characterize the prototypes developed by advanced L2 learners, at least for sounds differing along an acoustic dimension such as VOT.

The perception of certain dimensions in prototypes may be more resistant to change than certain others. One recent study suggested that certain stimulus dimensions will be more salient to learners than others. For example, Flege and Hillenbrand (1985) found that L2 learners were considerably more sensitive to variations in vowel than fricative duration as perceptual cues to the contrast between English /s/ and /z/ in a “piece” to “peas” continuum in which both vowel and fricative duration were varied in equal steps. Soli (1983) hypothesized that it is auditorily easier to perceive variations in vowel than fricative duration because of the relatively greater intensity of vowels. The lower frequency of vowels than fricatives might also be important.

Second, experience with L2 sounds might result in an “enriched” phonetic space, one in which similar L1 and L2 sounds have different, or at least functionally separate, prototypes. This is what we might expect of a “true” bilingual. This type of organization might contribute to the ability of an individual to perceive L1 and L2 sounds just like monolingual native speakers of L1 and L2. However, no study, to my knowledge, has demonstrated an instance in which subjects have perceived L1 and L2 sounds exactly like monolingual native speakers of their two languages.

One way to test for the existence of functionally separate L1 and L2 prototypes would be to determine if individuals who speak two languages perceive similar L1 and L2 sounds differently as a function of language. Evidence of perceptual “switching” has been found in only one of three experiments that have looked for it (Elman et al., 1977), and then only for the most advanced L2 learners.

It may be the case that perceptual switching is not instantaneous, but requires some time to occur, as well as the proper context. Evidence from speech production reported by Grosjean and Soares (1986) suggested that switching between languages cannot be accomplished easily for short stretches of speech. L1 words inserted into L2 sentences were found to be acoustically “tinged” by language context. This suggested the possibility that instead of choosing from a different array of prototypes, talkers reset broad production parameters when they switch between languages.

Third, phonetic category prototypes might be modified to accommodate the different acoustic properties of similar L1 and L2 sounds. If this were so, L2 learners would be expected to perceive both L1 and L2 sounds differently than monolingual speakers as the result of their experience in two languages. Williams (1980) noted that evidence of a “compromise” in speech perception might take two forms in experiments employing a synthetic continuum of sounds differing in a continuous parameter like VOT. The boundary between categories could occur at parameter values intermediate to those seen for monolingual speakers of L1 and L2, as was seen for the Puerto Rican children in Williams’ (1980) study. Or, there might be a relatively large number of stimuli which are labeled inconsistently, meaning that the “boundary” region is widened to embrace the category boundaries found in L1 and L2. This pattern, which would be shown by a more gradual slope in the identification function or by nonmonotonic functions, can be seen in child L1 studies (e.g., Zlatin & Koenigsknecht, 1975, 1976) and in the perception of L2 stops by French Canadians in the Caramazza et al. (1973) study.

Fourth, the L2 learner might abandon L1 category prototypes in favor of prototypes identical to those elaborated by native speakers of L2. This seems to have been true for some of Williams’ (1980) adult Spanish/English bilingual subjects (see also Mack, 1984). As suggested earlier, such a change might occur in instances where the phonetic contrast in L2 is easier to discriminate for auditory reasons than the L1 phonetic contrast.

Inasmuch as it is influenced by linguistic experience, the exact nature of prototypes developed for the production and perception of L2 speech sounds is apt to be affected by factors such as the age of acquisition, the nature of early learning and usage patterns, bilingual balance, and the little understood psychosocial factors mentioned earlier. It is unlikely that any of the hypotheses offered above will account uniquely for perceptual modification during L2 learning. However, before any of the above hypotheses can be eliminated or constrained more narrowly, we must first know more about the perception of L2 sounds by language learners.

**MECHANISMS CAUSING ACCENT**

**MAINTENANCE**

As noted earlier, L2 learners often seem to produce L2 words with L1 sounds. This happens whether the L1 word contains sounds that are acoustically identical to sounds found in L1, do not resemble any sound in L1 (i.e., new sounds), or are acoustically different from their closest counterpart in L1 (i.e., similar sounds).
This phenomenon is often referred to as "interference," a term deriving from paired associate and list learning experiments. "Proactive interference" refers specifically to a decrease in memory as the result of previous training or experience. In the context of L2 learning, it has been used to describe the effect of LI learning on L2 performance. The production of a sound is viewed as a motor "response" that has been "stamped in" or "overlearned" through repetition and reinforcement. The "stimulus" is either a sound that is perceived (as in an imitation task) or a sound category that is selected for production. In this view, the seeming advantage of early compared to late L2 learning might be due to less proactive interference for young children than adults which, itself, derives from weaker stimulus-response associations.

A term preferred over interference by many researchers is "transfer." "Positive" transfer is said to affect the production of similar or identical L2 sounds, for native speakers of L2 are unlikely to hear a sound substitution if a non-native replaces a sound in L2 with an identical or similar LI sound. Spanish /i/, for example, bears sufficient similarity to English /i/ that it would probably not be perceived as a substitute were it to replace English /i/. If so, transfer would be regarded as "positive" because the Spanish L2 learner has not needed to develop a new motor plan (or phonetic "goal") for realizing English /i/.

The same acoustic difference might legitimately be regarded as leading to negative transfer at another level of analysis. Even though use of Spanish /i/ in English words is unlikely to lead to the perception of a different category, it might lead native speakers of English to detect foreign accent in the speech of a Spanish learner of English (see Johansson, 1975; Flege & Hammond, 1982; Flege, 1984a), or slow the processing time needed for category identification or discrimination (Redmond, 1977).

The term "maintenance" may be more appropriate than either interference or transfer because of the broad similarities between L1 and L2 speech learning. Milisen (1954) speculated that motor skill development during a "pre-speech" period contributes later to the articulation of speech sounds of LI. Locke (1983a; see also MacNeilage et al., 1981) observed that the sounds most frequently noted in the babbling of pre-articulate infants are maintained in early speech. For example, one of the sounds most frequently heard in the pre-articulate babbling of babies in many language environments is a /d/-like sound, which occurs often in early speech.

S. Fletcher (personal communication) suggested that the high frequency of alveolar (or palato-alveolar) sounds like /d/ in babbling and early speech may be due to the baby's continued use of a reflexive tongue gesture employed for suckling. Because of infants' early establishment of control for a tongue closing gesture that leads to a horseshoe-shaped occlusion of the oral cavity, /d/ may be used by the child to produce other sounds not yet controlled. This hypothesis is consistent with the frequent observation of child L1 acquisition studies showing that children often replace developmentally early-occurring sounds for later-occurring sounds. A repercussion of the seeming early preference for /d/ seems to be its high frequency of occurrence in the English lexicon and in conversational speech.

As in L1 acquisition, a basic mechanism in L2 speech learning seems to be the maintenance of previously established patterns of articulation for the production of L2 sounds that bear varying degrees of acoustic and articulatory similarity to LI sounds. This can be illustrated by considering the production of English /r/ and /l/ by native speakers of Japanese. Hay and Fukuzawa (1979) analyzed spectrographically productions of /r/ in Japanese words. This "dral" sound resembles a flapped English /d/ in some phonetic contexts. They found that the Japanese /r/ resembled English /l/ acoustically to a greater extent than English /r/. Although native speakers of English heard /l/-for-/r/, and to a lesser extent /r/-for-/l/, substitutions in English words spoken by Japanese native speakers, the authors suggested that Japanese learners often do not substitute one English sound for another (e.g., English /l/ for /r/), but instead substitute the Japanese /r/ sound for the /l/ and /l/ sounds of English. If correct, this would tend to confirm Valdman's (1976) claim that L2 learners maintain LI patterns of articulation for the production of L2 sounds.

Additional evidence for this procrustean solution to the problem of producing new or different L2 sounds is provided by Cochrane (1977). She trained listeners to transcribe the /r/ and /l/ in English words spoken by Japanese L2 learners as omitted, substituted for one another (e.g., /l/-for-/l/), or substituted by another sound of English (e.g., /d/ or /w/), or substituted by Japanese /r/. Adult Japanese subjects used the Japanese /r/ about 17% of the time, children only 3% of the time. This suggests the possibility that the adults tended to maintain previously established patterns of articulation to a greater extent than the children.

A number of instrumental phonetic studies examining details of segmental articulation, to be reviewed later in this chapter, have also indicated that LI patterns of phonetic implementation are regularly maintained in L2 production.

**LEARNING**

It is likely that not all of the sounds needed by the child to produce L1 words have occurred spontaneously in babbling or derive from innate or reflexive oral motor control patterns. If so, the L1-learning
child—like the L2 learner—must learn to produce a variety of new sounds, the number and nature of which will depend on the phonetic inventory of the L1 being acquired. We know that thorough mastery of certain sounds takes longer than that of other sounds. But children possess some kind of error correction mechanism allowing them to compare the results of each attempt to their developing notion of how each sound ought to be produced. As a result, children's "replica" sounds gradually approximate adults' "model" sounds until the child finally produces L1 sounds like mature native speakers.

One weakness of the segmentally-based approach to L2 learning known as Contrastive Analysis (Flege, 1979; Mitleb, 1981 for discussions) is that it implies all-or-nothing success or failure in producing the sounds of L2. In some instances L2 learners have been observed to produce an L2 sound with a single L1 sound, but in other instances they have been observed to use a wide range of variants in attempting to produce an L2 sound. The general assumption (Haugen, 1950, p. 285) is that, although the learner may initially vacillate in the choice of which LI sound to substitute for a sound in L2, he or she will finally settle on an L1 substitute sound to use.

A weakness of approaches based on a static, cross-language comparisons of sounds (or phonemes) in L1 and L2 is that they are not able to generate predictions concerning the learning which occurs or fails to occur as the L2 learner gradually approximates L2 phonetic norms. This weakness is crucial, since many studies have shown that children and adults do learn, with varying degrees of success, to produce new or similar L2 sounds. For example, Dickerson (1974, 1975) transcribed the English sounds produced by native speakers of Japanese, arranging the different variants used to produce a number of English sounds (/r,l,s,z,θ/) according to how closely they approximated English phonetic norms. Over the course of a year in the United States, the subjects produced increasingly fewer variants judged to be distant approximations, and more variants judged to be close approximations.

A number of cross-sectional instrumental studies to be reviewed later in this chapter have suggested that experienced L2 speakers often approximate L2 phonetic norms more closely than relatively inexperienced learners. For example, Flege (1980) presented individual data for two groups of Saudi learners of English. The experienced subjects approximated the phonetic norms for English stop production more closely than the less experienced subjects, but almost all of the subjects produced stops with acoustic properties that were intermediate to those typical for Arabic and English.

The existence of phonetic approximation suggests that the L2 learner has noticed, and is responding to, differences between sounds in the

L1 and L2 phonetic inventory. In recent years learning has most often been described as the result of changes in the learner's "interlanguage" phonology. Based on his observations of a Hungarian learner of English, Nemser (1971a, 1971b; see also F. Johannson, 1973) concluded that variability in the production of an L2 sound is the product of a "linguistic system," rather than just random variability. Corder (1967, 1971) concluded that systematic departures from the phonetic norms of L2 are the result of an "interlanguage phonology." Thus when the L2 learner produces a sound not heard on the phonetic surface of LI or L2, it should not be viewed as a failure or sign of incompetence, but as evidence of incomplete learning. Less systematic errors, on the other hand, may not necessarily provide evidence that learning is underway although they, too, may demonstrate awareness of differences between LI and L2 sounds.

Three major kinds of evidence have been presented to support the existence of an interlanguage phonology distinct from the phonology of LI or L2. The first is the existence of pidgins and creoles, which may be regarded as linguistic systems that blend the properties of two languages. The second is the observation that L2 learners tend to make different pronunciation mistakes as L2 learning progresses, and that pronunciation of L2 target sounds often evolves towards an increasingly close approximation of L2 phonetic norms (Dickerson, 1974; Flege, 1980).

The third kind of evidence is the observation (e.g., Nemser, 1971a) that some L2 sounds produced by the L2 learner are not to be found on the phonetic surface of either LI or L2. It is likely that, in these instances, the learner is not using an LI sound without at least some modification. Moreover, it suggests that when the learner sporadically produces an L2 sound authentically it does not mean necessarily that he or she has developed a central phonetic representation which is identical to the one used by native speakers of L2 to produce that sound.

An interlanguage phonology, as described by a number of researchers, differs from the phonology of LI and L2 in that it is actively constructed by the L2 learner and evolves during the course of L2 learning. Selinker (1972) proposed that the learner makes generalizations about L2 sounds and their structural relationships to one another based on experience speaking and hearing L2. As the result of this experience, the learner's interlanguage gradually approximates the phonology of L2 native speakers, although it may ultimately reach a limit beyond which further experience does not lead to a closer approximation of the phonology of L2 native speakers. Selinker suggested that such "fossilized" inter-
language phonologies should be referred to as “idiolects” because of their resistance to further change. It is unclear why fossilization occurs.

An implicit assumption behind the notion of interlanguage phonology is that the L2 pronunciation of learners with the same L1 will evolve in a similar fashion. Nemser (1971a,b) suggested that individual differences in L2 production are likely to derive from differences in L2 input or experience. A production study by Flege (1980) revealed substantial intersubject variability as defined by the standard deviation of replicate measures of VOT, vowel duration, and stop closure duration. The source of this variability is not clear. One possibility is that the concept of “interlanguage phonology” does not apply to phonetic implementation or realization. Another possibility is that some subjects had quantitatively or qualitatively better phonetic input than others.

Still another possibility is that differences in L2 speech production reflect the same kind of non-linear development seen in L1 acquisition. Flege and Davidian (1985) noted that adult L2 learners manifested two tendencies seen in the speech of children learning English as an L1. Chinese and Spanish learners whose L1 contained no word-final stops tended to delete stops in the final position of English words. However, Polish learners of English (whose L1 does contain word-final stops) did not delete stops, even including /b,d,g/ (which are not found in the final position of Polish words). All of the non-native speakers (Poles, Chinese, and Mexicans) tended to devoice word-final /b,d,g/ (see Eckman, 1977, for a relevant discussion of markedness).

One possible interpretation of these data is that the developing ability of the non-native speaker to produce voiced word-final stops in English resulted from learning to suppress the “natural” processes of final stop deletion and final stop devoicing (see Stampe, 1979). This interpretation is consistent with the observation that English-learning children who are not exposed to a devoicing rule in their L1 also devoice /b,d,g/ (see Flege, 1982), and that Chinese speakers not exposed to word-final stops in their L1 also devoiced final /b,d,g/ in the Flege and Davidian (1985) study.

Wode (1977) provided an illustration of what appears to be a natural process in L2 learning by children. English-learning children often substituted /w/ for the initial sound in words like “rabbit.” Wode observed that three out of four German children substituted the motorically “easy” uvular /r/ of German for the more difficult English /r/ (see McLaughlin, 1978, p. 72), and substituted a /w/-like sound for English /r/. This last substitution pattern was clearly not a case of the maintenance of an L1 sound in L2, for German has no /w/. It probably resulted from an awareness that German and English /r/ differ. That is, the partial articulatory and visual similarity between English /w/ and /r/ which leads the English-learning child to substitute /w/ for /r/ probably also caused the German children to do the same even though they were somewhat older (4 to 8 years of age) than many of children who do so when acquiring English as an L1.

Because an interlanguage phonology is thought to be “internally structured” and “cohesive,” some believe that it can be studied like the phonology of natural human languages (Tarone, 1980, 1981; Eckman, 1981). Others, however, argue that it is not possible to study interlanguage phonology because L2 pronunciation is too sporadic and changes too rapidly, and because individual differences are too great (Slama-Cazacu, 1971).

Most of the work on interlanguage phonology has been largely theoretical rather than observational or experimental. As a result, the existing literature neither supports nor refutes the hypothesis that learners develop a separate phonology for the production of L2. (For a discussion of single versus merged language systems in bilinguals, see Albert and Obler, 1978.) Nor does the evidence now available provide a clear idea of whether L2 learners develop new or functionally separate phonetic category prototypes for similar and new sounds in L2.

LOSS

It has been observed frequently that infants babble sounds not heard in the speech of the surrounding community but that, in time, they cease producing sounds which differ from those found in the phonetic inventory of the L1 they are acquiring. Locke (1983a) observed that certain sounds are gradually “lost” from the infant’s repertoire of babbled sounds. For example, /h/ may constitute over half the consonantal sounds produced by a babbling infant, but its frequency is reduced to the same level as in adult speech (about 6%) by about two years of age. One broad hypothesis is that when a child begins to speak L1, he or she makes increasing use of sounds that have been selected, and decreasing use of sounds that are not under active, volitional control.

It is not certain whether volitionally selected sounds must be practiced in order for motor control to be maintained. Perhaps the information in phonetic category prototypes must be updated. If so, an L2 learner who stops using an L1 sound may lose some control of it if he or she stops speaking L1. This hypothesis could be tested by examining the production of an L1 sound that does not occur in L2 by an individual who speaks only L2. For example, a Xhosa native speaker who lives in an English-speaking community without any opportunity to speak or hear Xhosa should produce Xhosa sounds that differ from any sound
in English (e.g., clicks) less authentically than Xhosa speakers who routinely speak Xhosa, but pronounce Xhosa sounds that are identical to a sound in English with complete authenticity.

HYPERCORRECTION

There may be a discrepancy between how talkers think a sound “ought” to be produced (i.e., their perception of “norm”), and how they actually produce that sound in spontaneous speech. The difference between actual and idealized speech may lead to a phenomenon known as “hypercorrection.” For example, native speakers of /r/-less English dialects sometimes insert an /r/ into words that do not contain /r/ in the speech of other native speakers of English (e.g., “Cubar” for “Cuba”). They hypercorrect as the result of having observed that others sometimes produce an /r/ where they do not, and from the mistaken conclusion that certain words (e.g., “Cuba”) are lexicalized with /r/.

Hypercorrection also seems to affect L2 speech production. For example, native English speakers are apt to comment on the substitution of /b/ for /p/ (e.g., “barking lot” for “parking lot”) in the English spoken by Arabs. Because of this, many native speakers of Arabic have explicit knowledge of their difficulty in producing English /p/. Perhaps as a result of this knowledge, and because of uncertainty concerning which words in English are lexicalized with /p/, Arabs sometimes produce words like “barking” with a /p/ even though their L1 has no /p/ sound (B. Vann, personal communication; Flege, 1979).

Marcokwadt (1946; but cf. Michaels, 1974) observed that beginning Spanish learners of English produced English words ending in labial, alveolar, or velar nasal consonants (e.g., “sum,” “sun,” “sung”) with /n/, whereas more advanced Spanish learners produced all three words with final /n/. The substitution pattern seen in early stages of learning may be the result of maintenance. In certain Spanish dialects words end in /n/ but not /n/ or /m/. However, the substitution pattern observed in later stages of development may represent hypercorrection, that is, the overuse of a sound known to be difficult. Another possibility is that this phenomenon represents transfer of a casual speech variant, for /n/ is realized as /n/ in certain dialects of Spanish in casual speech.

It would be useful to test directly the hypothesis that L2 learners hypercorrect and, if so, whether it results from incorrect lexicalization or from a strategy for speaking L2 “correctly.” If hypercorrection is the result of an output strategy, one would expect to observe a lower frequency of hypercorrect productions (e.g., /p/ for /b/ in Arabic-accented English) in noncareful or emotionally-charged speech than in careful speech, and a greater frequency in a reading task (where the correct phoneme is orthographically represented) than in an imitation task.

LINGUISTIC FACTORS

A number of other factors which might be loosely termed “linguistic” also influence L2 pronunciation.

Orthographic Effects.

Literate native speakers of languages like Spanish and Finnish learn to associate orthographic symbols with specific sounds to a greater extent than speakers of a language like English, in which there is a less direct relationship between sounds and graphic symbols. English spelling leads to pronunciation difficulties for native speakers of “phonemic” languages, especially if their exposure to English comes principally through the written word, as it does in some formal programs of instruction. For example, native speakers of Spanish often produce English words like “big” as /big/ because the graphic symbol “i” is always used in Spanish to represent an /i/-quality vowel. Another spelling-based mispronunciation typical of Spanish native speakers stems from a lack of one-to-one correspondence in L1 rather than L2. Native Spanish speakers often pronounce words like English “vase” as /ves/ because orthographic “v” in Spanish is realized as /b/ or /v/ in Spanish words depending on phonetic environment (Hammond, 1982; Flege & Hammond, 1982). (Another possibility, of course, is that Spanish learners do not perceive the phonetic difference between English /b/ and /v/.)

The effect of orthography is clearly evident in problems related to vowel reduction. An important phonological characteristic of English is that stressed syllables tend to alternate with unstressed syllables. An important phonetic characteristic of English is that vowels in unstressed syllables tend to be “reduced”, that is, produced with vowel formants which places them nearer the center of the acoustic phonetic vowel space. Vowels in stressed syllables, on the other hand, tend to be produced as full vowels. In the word “substitution,” for example, the first “u” is realized as a reduced vowel whereas the second “u” is realized as a full /u/-quality vowel. Non-native speakers of English often fail to reduce unstressed vowels if their L1 differs from English in this regard. Spanish learners, for example, tend to realize both occurrences of “u” in “substitution” as /u/, and to realize the “i” as /i/, just as if it were found in a Spanish word.
Two unpublished studies cited by Walz (1979, p. 33) indicated that pronunciation may be less authentic when cued by orthography than by other means. However, I know of no empirical study which has assessed rigorously orthographic effects on L2 pronunciation. This could be done by comparing the English pronunciation of native speakers of “phonemic” (e.g., Spanish) and “non-phonemic” languages (e.g. Dutch), or by comparing the English spoken by individuals whose primary L2 experience was based on orthography (which is typical for many formal programs of classroom instruction) to the English spoken by individuals who learned L2 naturallyistically. Both conversational speech and more formal speech (e.g., paragraph and word-list reading) should be examined, ideally at several stages of learning.

If orthography influences L2 pronunciation, spelling pronunciations should be observed more frequently in the speech of native speakers of “phonemic” than “non-phonemic” languages, and for individuals who learned L2 formally than naturallyistically. These effects, if observed, should be greater for speech samples elicited aurally (e.g., in an imitation task) than orthographically, and in formal than casual speech.

An experimental approach to this question would be useful to control for potential confounding factors. For example, native speakers of languages with comparable vowel systems such as Spanish and Japanese could be trained to recall balanced lists of CVCVC nonsense words spoken by a native English speaker with stress on the first vowel. Ten-member word lists could be created in which the vowel in the first syllable (spelled with “i,” “e,” “u,” “a,” “o”) is realized as a full vowel in half the words, and as a reduced vowel in the other half. Here is a partial example of such a word list:

```
1st Vowel  2nd Vowel

"pokap" /o/     /̃/ 
"polap" /u/     /̃/ 
"musat" /i/     /̃/ 
"mulat" /i/     /̃/ 
```

To test the hypothesis, it would be necessary to assign randomly subjects to one of two treatment conditions in which the mode of stimulus presentation differs. In one condition, subjects would be trained to recall 10-word lists by means of the simultaneous written and aural presentation of the list items. In the other condition, subjects would receive only aural presentation of the items. The subjects would be asked periodically to name as many items from the list just presented as possible. The dependent variable would not be how many items are remembered, but what percentage of reduced vowels are reproduced as full vowels. If orthography affects how lexical items are pronounced,

...subjects in the first (written + aural) condition should reproduce reduced vowels as full vowels more frequently than subjects assigned to the second (aural only) condition. If subjects use orthographic representations to store words in memory, the effect should become stronger as subjects learn more items on a list. Subjects whose L1 is a phonemic language (e.g., Spanish or Finnish) might be expected to make greater use of orthography for the storage and retrieval of lexical items than speakers of “non-phonemic” languages. If so, they should produce reduced vowels as full vowels more often than subjects whose L1 is a non-phonemic language with no alphabetic representation of vowel sounds (e.g., Japanese).

**Word Familiarity.**

Thus far we have considered speech sounds as if they were produced in isolation. However, it has long been known that an individual’s familiarity with a particular lexical item influences how the sounds contained in it are produced. For example, a classic study by Leonard (1972) examined the effect of prior experience on phonetic learning. Ten English-speaking children who typically substituted /r/ for /l/ were given training with feedback on the correct production of /r/ in high-frequency English words such as “run” and “read” and matched nonsense words such as “ruz” and “reen.” The children took no longer to reach criterion (three consecutive productions of /r/ judged to be correct) for the real than the unfamiliar nonsense words. However, they took significantly longer to re-achieve criterion for the real words than the nonsense words in a second session 90 min later. Leonard speculated that articulatory patterns established previously by the children for the production of the real words proactively “interfered” with learning that occurred during the training session.

The importance of how many times a word has been heard previously was shown indirectly by Leonard and Ritterman (1971). They found that English-speaking 7-year-olds produced /s/ better in consonant clusters which occur frequently in English than in relatively infrequent clusters. When cluster frequency was held constant, they produced /s/ better in frequently than infrequently occurring English words.

If word familiarity affects L2 learning as it does L1 learning, L2 sounds may be produced less authentically in L2 words which have a cognate in the learner’s L1. “Cognate” words are words in L1 and L2 that are spelled alike and mean approximately the same thing, such as “nation” in French and English. Despite the fact that this cognate pair is spelled the same in both languages, they are distinguished by a number of phonetic differences. For example, the French form is pro-
duced with approximately equal stress on both syllables (albeit with slightly longer duration on the second syllable) and ends in a nasalized vowel. The English form ends in /n/ and has a diphthong (/eɪ/) rather than a monophthong (/a/) in the first syllable.

The pronunciation of cognate versus noncognate words in L1 and L2 has been the object of considerable speculation but little systematic inquiry. If patterns of articulation established for entire L1 words (not just sounds or syllables) are maintained in L2, sounds should in general be produced more authentically in L2 words that do not have an L1 cognate than in words with an L1 cognate provided that other variables such as phonetic context, degree of stress, and amount of experience are held constant.

Support for this hypothesis comes from Hammerly (1976, cited by Walz, 1979), who found four times as many pronunciation errors in cognate than noncognate words. Similarly, in a picture-naming task, Cochrane (1977) found that adult Japanese learners of English produced sounds less authentically in English words with cognates than in English words without cognate in Japanese. Japanese children, however, did not show the same difference (but cf. Winitz & Bellerose, 1965).

Several possible explanations for this last finding come to mind. The children may not have recognized the relationship between cognate words in L1 and L2. Or, children's articulation of L2 sounds may not be influenced to the same extent by the word in which it occurs as it does for adults (see Kent, 1983; Massaro, 1975; Crowder, 1976). Still another possibility is that proactive interference exerts less influence on children's than adults' pronunciation because children's articulatory "habits" are less thoroughly established. Whatever the explanation, the finding of Cochrane (1977) suggests a potential advantage for child compared to adult L2 learners that needs to be investigated further.

Word familiarity, which itself is related to frequency of occurrence, is closely related to the cognate versus noncognate status of words in L2. Cognate words are by definition "familiar" to the L2 learner the first time they are encountered in L2. Several unpublished studies cited by Walz (1979, p. 34) suggested that familiar L2 words are pronounced better than unfamiliar words. However, I know of no systematic empirical study which has examined the effect of word familiarity on L2 pronunciation. Nor has any systematic study using auditory appraisal of pronunciation or instrumental measurement examined speech sound articulation as a function of word familiarity or the cognate versus noncognate status of L2 words.

The two-stage model of speech production outlined earlier posited that category selection and sequencing occur independently of phonetic implementation and realization. A stimulus-response association model (Leonard, 1972; Winitz & Bellerose, 1972, 1975, 1978) highlights the potential importance of proactive interference. Both models predict an important difference between cognate and noncognate words in L2. For example, patterns of articulation established for English /t/ might be expected to exert greater proactive interference for /t/ in cognate words because once a word is recognized, processing of its constituent sounds can be expected to cease (Massaro, 1975).

If the sounds in a cognate word are processed less thoroughly than sounds in a noncognate word, proactive interference might be less important. Native speakers of English, for example, should produce the /t/ of French less authentically in cognate than noncognate words. More specifically, they should produce French /t/ with VOT values that are significantly longer (i.e., English-like) in French words with an English cognate, such as "tonique," than similar French words without cognates in English, such as "tonneau." Future research which controls both the familiarity and the cognate versus noncognate status of L2 words will be needed to test this hypothesis.

Phonetic Context.

Should the study just outlined be carried out, phonetic context and word position will also need to be controlled. This is because the authenticity of a sound as well as its articulation is likely to be affected by these factors (see Kent, 1982). Research has consistently shown that adjacent sounds are coarticulated to some degree. The multiple, covarying gestures used to produce a sound do not occur in discrete bundles, as suggested by distinctive feature theory or this chapter's use of the term "speech sound." Instead, they overlap and co-occur in time with the gestures used to produce neighboring sounds. Consider, for example, the tongue movement needed to make lingual-alveolar contact for /t/ when it is preceded and followed by /s/ in the phrase "the cast sank" compared to the /t/ in "the cot ought."

Several L2 studies have demonstrated that authenticity of pronunciation varies as a function of phonetic context. Gatbonton (1975, discussed in Segalowitz and Gatbonton, 1977) observed that the difficult /s/ sound of English was produced authentically by French Canadians more often when it preceded a vowel or fricative than a voiceless stop. Dickerson (1974, 1975) found that the perceived authenticity with which Japanese learners produced English sounds varied as a function of phonetic context. In both studies, the effect of phonetic context on authenticity persisted over time even though pronunciation improved overall.

In an unpublished study, Turitz (1981) examined the pronunciation
of English /s/ and /z/ in an English text read by 11 native Spanish speakers representing a wide range of English-speaking abilities. Word position and context appeared to affect importantly the frequency with which /s/ and /z/ were judged to have been produced "correctly":

<table>
<thead>
<tr>
<th>context:</th>
<th>#sV</th>
<th>#sC</th>
<th>VsV</th>
<th>Vs#</th>
<th>Cs#</th>
</tr>
</thead>
<tbody>
<tr>
<td>% correct:</td>
<td>97%</td>
<td>87%</td>
<td>96%</td>
<td>81%</td>
<td>56%</td>
</tr>
</tbody>
</table>

The subjects realized the two fricatives correctly less often in word-final than word-initial or intervocalic position, especially when they followed another consonant. This is not the pattern observed ordinarily in English L1 acquisition, where children often succeed in producing word-final fricatives before mastering word-initial fricatives (Mazza, Shuckers, & Daniloff, 1979). The word position effect evident in Turitz's data appears to be due to the maintenance in L2 of L1 patterns of articulation. Spanish /s/ is articulated many different ways in the Venezuelan dialects of Spanish spoken by her subjects. In some dialects it is realized as a sibilant in word-final position, in other dialects as an /h/-like sound, and in still others it is often omitted.

Since many Latin Americans are accustomed to the omission of final /s/, it might seem to them unimportant to produce /s/ at the end of English words. It would be interesting to examine the production of English final /s/ by just those Venezuelans who produce an /s/ in final position in their native dialect of Spanish. For them, maintenance of L1 patterns should result in positive transfer, so that no clear difference due to word-position would be expected. For native speakers of /s/-omitting dialects, on the other hand, we would expect the pattern noted above. Only with sufficient learning would the word-position effect be expected to disappear.

It should be noted, however, that sociolinguistic factors might delay learning or render complete learning impossible, and that attentional factors might also be important. For example, Turitz (1981) found that /s/ and /z/ were judged to be distorted or omitted more frequently in the second than first half of the reading passage examined.

Coarticulation.

Languages may differ according to the extent to which the articulation of a sound is permitted to influence the articulation of adjacent sounds. To the extent that coarticulatory patterns are learned in L1 acquisition, they are likely to be maintained in L2 production and lead to the perception of accent. However, to my knowledge no L2 research has specifically tested this hypothesis. The following discussion is therefore meant to suggest ways in which coarticulation might be examined in L2 research.

Lubker and Gay (1982) examined anticipatory labial coarticulation in the speech of native speakers of English and Swedish. They found that, compared to native English speakers, Swedes protruded the upper lip to a greater extent, showed less variability in the extent of upper lip displacement, and initiated lip protrusion earlier with respect to the acoustic onset of a following rounded vowel. The authors concluded (1982, p. 444) that native speakers of Swedish learn "different targets or goals for labial protrusion" than native speakers of English, and suggested that the difference between native speakers of Swedish and English was perceptually motivated. The Swedish vowel space is more crowded than that of English, with many more rounded vowels. Variations in the degree of lip protrusion is more likely to lead to confusions between Swedish vowels than among English vowels because the amount of lip protrusion affects the length of the lip orifice which, in turn, affects oral cavity length and thus the frequency of vowel formants.

A basic principle of speech communication is that talkers tend to produce reliably only those distinctions with perceptual significance. Were English learners of Swedish to maintain the English pattern of labial protrusion in producing Swedish rounded vowels, it might be expected to result in perceptual confusions and diminished authenticity. The maintenance of Swedish-like patterns of labial protrusion in English vowels, on the other hand, would be likely to lead only to diminished authenticity of vowel production. These considerations lead to the prediction that L1 patterns of labial coarticulation will be modified to a greater extent by English learners of Swedish than by Swedish learners of English.

Swedish and English words spoken by highly proficient Swedish learners of English and English learners of Swedish could be examined to test this prediction. If it were found that there is no difference in labial protrusion between Swedish and English words produced by any of the subjects, it would suggest that coarticulation is not modified during L2 acquisition. However, if just the Swedes modify labial protrusion when switching from L1 to L2, it would support the hypothesis that patterns of coarticulation are modified only when "communicative pressure" is brought to bear.

The hypothesis that L2 learners modify only those coarticulatory patterns with perceptual significance could also be tested by examining velar coarticulation. Proficient native speakers of English typically nasalize the vowel in words like "Tom" by lowering the velum before labial constriction of the word-final nasal consonant (/m/). The allo-
Phonetic nasalization of vowels is permitted in English because nasalization is not linguistically distinctive. In French, on the other hand, nasal and oral vowels contrast phonemically before non-nasal consonants like /b/ or /s/, and in word-final position. Before nasal consonants like /m/ and /n/ in French, however, only oral (i.e., non-nasal) vowels occur.

This difference in phonological structure leads to measurable differences in velar coarticulation. Clumeck (1976) reported that English speakers began lowering the velum sooner (an average of 186 ms) before the oral constriction of a following nasal consonant than native speakers of French (114 ms). Native speakers of French might delay lowering the velum when producing oral vowel + nasal consonant sequences because, in French, nasalization cues the perception of a nasal vowel (see Beddor & Strange, 1982). It appears that English native speakers, on the other hand, learn to lower the velum long before constriction of word-final nasal consonants because allophonic nasalization is not an important cue to vowel category identity in English (Thompson & Hixon, 1979).

These differences between French and English permit a second test of the hypothesis outlined earlier. The prediction is that French learners of English will be less likely to modify L1 (French) velar timing patterns when speaking L2 (English) than English learners of French L2. More specifically, it is predicted that French learners of English will lower the velum at approximately the same time before constriction of /m/ in comparable French and English words. This is because a French-like absence of allophonic nasalization in English words will not affect adequacy of pronunciation. English learners of French, on the other hand, are predicted to lower the velum later (i.e., closer in time with respect to the constriction of /m/) in French than English words because the failure to do so might lead to the perception of nasal vowels. These predictions again follow from the assumption that L1 patterns of coarticulation will be modified more often, or to a greater extent, when the failure to do so has important perceptual repercussion.

**Basis of Articulation.**

Phoneticians have detailed many specific differences in segmental articulation between languages. Some researchers believe that many of the specific differences which have been noted stem from a more subtle yet pervasive underlying difference between languages referred to as their “basis of articulation” (or “articulatory setting”). It has been proposed that the basis of articulation of a language is important to the “fine-tuning” needed to speak it without accent, even though the basis of articulation might not be revealed in segmental phonetic transcriptions. Chomsky and Halle (1968, p. 295) refered to it as something that confers on a language “its general phonetic aspect.” Honikman (1964) asserted that L2 learners cannot produce L2 with complete authenticity when the basis of articulation of L1 and L2 differs unless he or she changes the bases of articulation when switching from L1 to L2.

In discussions of basis of articulation, reference is often made to differences in phonetic inventory. Such differences may reveal a preference for certain types of sounds brought about by a language’s basis of articulation (or possibly the reverse). In some languages, for example, there is a preponderance of sounds produced at the front or rear of the oral cavity (see Maddieson, 1984). The phonetic inventory of sounds in various languages may differ according to overall degree of constriction, perhaps due to the existence of many palatalized consonants and close vowels. Reference is frequently made to cross-language differences in the neutral “rest position” the tongue assumes just prior to articulation. The primary evidence that tongue rest positions differ across languages is the auditory-based observation that the hesitation vowels produced during pauses by speakers of different languages may vary (Ozga, 1976).

Discussion has often focused on differences that are visually evident. Researchers have observed that a difference in basis of articulation may lead speakers of different languages to “look” different when speaking (Honikman, 1964). For example, some learners of English complain that native English speakers do not “open their mouths” when speaking, so that interincisal distances are relatively small and the tongue rarely seen. Ozga (1976) suggested that this may be due to the fact that open vowels constitute a much smaller percentage of vowels in English (24%) than in other languages like Polish (69%). He suggested that Poles learn to move the mandible more because their L1 requires the production not only of vowels differing greatly in height, but also a large number of palatalized consonants. Ozga also noted that the tongue tip is “anchored” to the lower incisors during the production of sounds in Polish, and the tongue body assumes a convex shape. In English, on the other hand, the primary anchorage is said to be between the lateral tongue margins and the molar teeth, and the basic shape of the tongue concave.

Delattre (1953) postulated that three general properties underly many discrete articulatory differences between the sounds of French and English. First, French sounds tend to be articulated with gestures that are more “tense” and energetic than English sounds. This general property was said to underlie the nondiphthongal quality of French
vowels, the lack of affricates in French, and the relatively small variation in intensity and fundamental frequency distinguishing syllables in a French word. Second, French sounds have a generally more anterior place of articulation than English sounds. As a result, French has more front and peripheral vowels than certain other languages, the anchoring of the tongue is behind the lower incisors in vowel production, French has a dental as opposed to alveolar place of articulation for /t,d,s,z/, and there is a relatively great degree of labialization of consonants preceding rounded vowels. Finally, Delattre described French syllables as being initiated less abruptly, and ending more sharply, than English syllables. This property was said to explain why French has more consonant-vowel syllables than English, the strong tendency of French consonants to be coarticulated with a following vowel (even across word boundaries), and the relatively strong tendency for French stops to be produced with an audible release burst (even to the extent of inserting of a schwa-like vowel after word-final stops).

Drachman (1970) speculated that young children may share a common basis of articulation that begins to disappear early in L1 acquisition. This hypothesis is supported by de Boysson-Bardies and Sagart (1983), who observed that the effect of different overall adjustments of laryngeal and supralaryngeal articulators between languages was evident in the speech of infants as young as ten months of age. Their analyses of short and long-term energy spectra of syllables produced by three French, Arabic, English, and Chinese infants was said to reveal “pervasive differences” in the quality of the attack and release phases of syllables, in voice quality, and in prosodic dimensions such as rhythm and intonation. Chinese babies, like Chinese-speaking adults, were observed to produce short syllables with creaky voice and abrupt offset of syllable energy. Arab babies, like native Arabic adults, were observed to produce final stops with a “progressive and turbulent release.”

Ozga (1976) noted that speculations about basis of articulation in the late nineteenth century provided the first real attempt at contrastive analysis. However, although some researchers claim that basis of articulation can be taught easily and contributes importantly to authenticity of L2 pronunciation, the “vagueness . . . and random selection of parameters in the description of bases of articulation” has prevented it from being seriously investigated in L2 research (1976, p. 65).

One exception to this is Homiedan’s (1984) study of tongue anchorage. Homiedan objectively tested the hypothesis that whereas native English speakers anchor the lateral margins of the tongue more against the upper than lower molars, native speakers of Arabic show the reverse tendency. The hypothesis, which was based on introspective judgments by native speakers of Arabic and English, seemed plausible in view of the existence of many pharyngeal consonants in Arabic but not English. Using the dynamic palatometry system of the Biocommunication Research Laboratory in Birmingham, Homiedan calculated the percentage of time tongue contact was maintained against the upper and lower molars in the production of sounds, words such as “bat”, and phonetically balanced Arabic and English sentences. It was found that, in sentence production, native speakers of Arabic maintained tongue contact against the lower molars in a significantly larger percentage of 10-ms sampling frames (about 98%) than native English speakers (90%). The English speakers maintained upper molar contact a somewhat larger percentage of the time (96%) than the native Arabic speakers (89%).

Given the general lack of empirical evidence, we can reach no firm conclusions regarding the importance of basis of articulation to L2 learning. However, we can sketch several experiments designed to test the hypothesis that basis of articulation influences L2 speech production. First, if L1 and L2 differ in basis of articulation to the extent that monolingual native speakers “look” different when talking, native speakers of L2 should be able to determine how authentically a learner produces L2 by looking at videotapes played back without sound.

Second, dynamic palatography could be used to assess the extent to which the overall vocal tract configuration is shifted in L2 production. As noted earlier, native speakers of French are said to produce sounds with a more “anterior” tongue placement than native speakers of English. If basis of articulation is related to foreign accent, then French learners judged to pronounce English authentically should produce /t,d,s,z/ with a more posterior tongue contact in English than French learners who are judged to pronounce English less authentically. Moreover, French learners with a good accent should show a greater change in tongue contact area in producing similar French and English sounds than French learners with a less authentic accent.

A third experiment could be directed at velar functioning. Clumeck (1976) noted that native speakers of American English but not French lowered the velum during production of the low vowel /a/, even when it was not adjacent to a nasal consonant. This suggested a difference in the “rest” position of the velum deriving from the existence of phonemically nasal vowels in French but not English. To the extent that L2 learners can alter a dimension like velar rest position, English learners of French should learn to suppress nasalization of /a/ in French words, and French learners of English should learn to nasalize English /a/.
In this section we will review research examining the production and perception of L2 speech sounds.

PHONOTACTICS AND SYLLABLE STRUCTURE

Sounds are not produced in isolation, but occur in sequences of syllables, words, and phrases. As a result, the pronunciation of a sound depends on factors that are extrinsic to its identity as a phonetic segment. The position of a sound within the syllable is undoubtedly important in L2 learning. For example, Brière, Campbell and Soemarmo (1968) reported that English speakers mispronounce L2 words beginning in /ŋ/ more often than words beginning in /ʒ/. This seems to follow from the fact that although both sounds occur in the phonetic surface of English, /ʒ/ but not /ŋ/ occurs in syllable-initial position (in words like /brɛŋs/, “brazier”).

Phonotactic constraints also play a role in determining how learners produce L2 sounds. Languages differ according to the sequences of sounds permitted to occur within a syllable. For example, the sequence /sk/ but not /ks/ is permitted in the initial position of English syllables. Such phonotactic “constraints” are often described with reference to syllable boundaries because sequences not permitted syllable-initially may occur across a syllable boundary. For example, “ksin” is not a possible English word because /ks/ is not permitted syllable-initially, yet this sequence does occur across a syllable boundary in the word “backstop”.

Some limitations on sound sequencing probably follow from physiological constraints such as the tendency to move from tighter to looser constrictions at the beginning of a syllable. However, many of what are referred to as phonotactic constraints are probably learned and language-specific, for consonant clusters not permitted in one language may occur in others. As is usually the case when differences in sound patterning exist across languages, learning difficulties arise for many L2 learners. For example, Spanish speakers often produce “school” as “eschool” because English, unlike Spanish, permits a /sk/ cluster in syllable-initial position. Broselow (in press,b) reported that, in attempting to produce unfamiliar L2 consonant clusters, learners have been observed to insert a vowel before the unfamiliar cluster, insert a vowel between members of the cluster, delete one member of a cluster, substitute one sound for another, and change the order of sounds in the cluster in order to produce a permissible sequence.

There are a number of factors that may lead to syllable structure pronunciation errors such as those just mentioned. It is possible that at least some of them are perceptual in origin. A talker may have the motor control needed for the implementation of two or more sounds in a rapid sequence, but not realize that the sequence consists of more than one sound. Awareness of which sounds constitute a word seems to be a metalinguistic skill that depends to some extent on experience with alphabets and writing (Morais, Cary, Alegria, & Bertelson, 1979). It is conceivable that L2 learners might not recognize the identity or order of sounds occurring in an unfamiliar L2 consonant cluster. Indirect support for this was provided by Warren, Obusek, Farmer, & Warren (1969), who found that subjects were unable to determine the temporal position of an unfamiliar auditory event (a cough superimposed on the speech signal) even though it was longer (>200 ms) than most speech sounds.

There is evidence that individuals have tacit knowledge of which sound sequences are permissible in their L1. Greenberg and Jenkins (1964) demonstrated the validity and reliability of judgments made by adult English speakers concerning the extent to which nonsense words conformed to the phonotactic constraints of English. Subjects’ evaluation of the acceptability of various CCVC syllables appeared to depend more closely on the sequencing than identity of sounds in the consonant cluster, for the presence of non-English sounds did not influence systematically any of their measures. Children also appear to be aware of the phonotactic constraints characterizing their L1. Findings similar to Greenberg and Jenkins’ (1964) were reported for English-speaking children aged 9 to 11 years by Pertz and Bever (1975). Messer (1967) found that three- and four-year-old English-speaking children chose a larger percentage of permissible than nonpermissible sequences when asked to indicate which of two forms was “more like a word.”

At present we have less certain information concerning whether L2 learners have the same tacit knowledge of L2 phonotactic constraints as L2 native speakers. One study (d’Anglejan, Lambert, & Tucker 1971) which examined French-speaking subjects’ knowledge of French phonotactics indirectly suggested that acquiring an L2 may influence judgments of phonotactic permissibility in L1. The study made use of the “linguistic distance” metric developed by Greenberg and Jenkins (1964) to evaluate French speakers’ subjective impression of the distance of various CCVC sequences from the norms of French. There was a correlation between an objective “distance from French” measure and subjects’ subjective ratings of the stimuli. However, there was great variability, including large differences in the subjective evaluations given to words assigned the same linguistic distance from French. The authors noted that intersubject variability may have resulted from varying
be more difficult to produce than familiar L1 sequences. It takes some time during L1 acquisition for children to master productive control of the full range of the syllable types permitted. Children often appear to master a sound earlier in one syllable or word position than another. Most children are said to "simplify" syllables (Ingram, 1976) at some point in L1 development. For example, they commonly reduce initial clusters and delete final consonants (e.g., produce "spot" as /ba/). Some children learning L1 have been observed to attempt only words which conform to a "preferred" syllable shape (Menn, 1971, 1980).

A finding of Messer (1967), however, suggests that children's syllable structure errors may be the result of the tendency to maintain only pre-established patterns of articulation when attempting to produce syllables not found in their phonetic repertoire, rather than use of an active strategy of syllable simplification. Like Menyuk (1968), Messer found that children correctly repeated words with permissible sound sequences more frequently than those with nonpermissible sequences. However, he noted that when children mispronounced nonpermissible clusters, they did so in such a way as to make them more nearly conform to the phonotactic constraints of English.

As observed earlier, the tendency to maintain established patterns of production may also characterize L2 acquisition. Tarone (1981) observed that L2 learners often experience difficulty producing syllables which do not occur in their L1. Broselow (in press) suggested that L2 learners attempt to "bring second language forms into conformity with first language restrictions" when faced with the need to produce unfamiliar sound sequences in L2. A study by Tarone (1980) supported this hypothesis. Tarone examined short excerpts of the English spoken by two native speakers each of Korean, Portuguese, and Cantonese. Cantonese and Portuguese permit few consonants in syllable-final position, whereas Korean more nearly resembles English in terms of the range of permissible syllable-final consonant clusters. From 16-24% of the syllables produced by the non-native subjects evidenced a syllable structure error. About 75% of the observed errors seemed to represent the replacement of an L2 syllable not found in L1 by a syllable found in L1.

This suggests that, like the child acquiring L1, the L2 learner also seems to rely on previously established (L1) syllables, at least during the initial stages of learning. However, as in L1 acquisition, maintenance is probably not the only factor affecting syllable structure errors in L2 production. If pronunciation errors were due entirely to maintenance, the L2 learner should mispronounce L2 syllables not occurring in L1, but never mispronounce syllables which do occur in L1. Tarone (1980)
observed that a small number (about 4%) of observed syllable structure errors could not be attributed to maintenance in any obvious way. For example, a Korean produced the word “whole” as /ho:/, even though a similar syllable (/ko:1/) exists in Korean.

Errors such as the one just mentioned might have resulted from a phonological selection error deriving from the learner’s failure to correctly lexicalize the target word. They might also be “performance” errors such as those which occur in L1 speech production. Since L2 learners must divide their attention over many domains (word choice, syntactic structure, pronunciation, and so on), it seems reasonable to think that L2 learners manifest as many “honest mistakes” as native speakers of L2. However, since Tarone did not provide data for an English control group, there is no way to evaluate this hypothesis.

“Ease of articulation” may interact with the language learner’s general strategy of relying on previously established articulatory patterns. Tarone (1981) suggested that when L2 learners must produce a syllable not found in L1, they tend to respond with CV syllables because this syllable type is easier to produce than any other. This conclusion concerning the ease of articulation of CV syllables is based on the frequent occurrence of CV syllables in infants’ reduplicated babbling and in human languages. A number of talkers in Tarone’s (1980) study manifested a tendency to produce complex syllables as the “universally” preferred CV syllable. The Korean and Cantonese talkers tended to do so by deleting a consonant more often (70%) than inserting a vowel into a cluster (30%), whereas the two Portuguese talkers inserted a vowel much more often (80%) than they deleted a consonant (20%). Both the consonant deletion and vowel insertion “strategies” for creating CV syllables occurred more often in syllable-final (40%) than syllable-initial position (11%).

One other source of error may not be directly related to either productive or perceptual aspects of speech processing, but pertain to the learned, phonological organization of languages. Broselow (in press, a) hypothesized that L2 learners perceive and produce L2 sound sequences according to L1 phonological rules. Based on the mispronunciation of Arabic words by native English speakers, she observed that native speakers of English tend to perceive incorrectly the location of word boundaries (denoted by the symbol #) in Arabic words. For example, /?ilkursi#gidiel/ is heard as /?ilkursig#idiel/, /mi#ana/ as /mi#jana/, and /binti#smiina/ as /bintis#miina/. Broselow asserted that such perceptual errors arise from the strong tendency for word boundaries to coincide with syllable boundaries in English but not Arabic. Since Arabic syllables usually begin with a single consonant, the first member of a CC cluster is linked to the preceding word in speech production. This leads to a syllable boundary between the two consonants. A native

English /r/ and /l/.

Many languages (e.g., Chinese and Korean) do not possess a contrast between /r/ and /l/. These two kinds of sounds are usually referred to as “liquids,” a name which suggests incomplete constriction. Maddieson
(1984) reported that about 40% of the 317 languages he surveyed had fewer than two liquid sounds. Most (about 77%) languages had some kind of /r/, usually a tongue-tip or uvular trill, a flap, or a tap produced with light contact between the tongue tip and alveolar ridge. Few languages (about 8%) had an /r/ resembling that of English. English /r/ is produced either by retroflexing the tongue tip, or lifting the mass of the tongue body towards the palate and simultaneously drawing the tip back into the body of the tongue.

The most common kind of /l/ reported in Maddieson's (1984) survey was a voiced lateral (about 75% of languages), usually formed with constriction in the alveolar region (about 80% of instances). English /l/ is formed by making lateral contact between the lateral margin(s) of the tongue and the teeth (some combination of incisors, canine, or pre-molars) and gingiva. In prevocalic position, the tongue dorsum approximates the palate, but in other positions a “dark” /l/ is formed by raising the tongue dorsum.

Many children learning English and other languages with two liquids have special difficulty mastering /r/ and/or /l/. Liquids are mastered later in L1 acquisition by normally developing children than many other sound types, and pose a special learning difficulty for children whose general speech development is delayed (Jakobson, 1968; Strange & Broen, 1980). Difficulty with /l/ and /r/ might be motivated by an inherent articulatory difficulty in producing liquid sounds which derives from the relatively small area of contact between the tongue and stationary articulators (the teeth, gingiva, and hard palate).

Difficulty in producing /r/ and /l/ might also be motivated by perceptual factors. An important acoustic difference distinguishing English /r/ and /l/ is the starting frequency of the third formant and its rate of transition into or out of an adjacent vowel sound (McGovern & Strange, 1977). F3 starts out much lower in frequency for /r/ than /l/, has a shorter steady state, and changes more slowly in frequency. A factor that may make the acquisition of liquids difficult is the wide range of acoustically different allophones for /l/ and, to an even greater extent, /r/. A relatively great amount of allophonic variation in the realization of a category probably has the effect of making acoustic cues (e.g., the starting frequency of F3) relatively difficult for learners to isolate and use in perceptual processing. This, in turn, might lead to an incorrect lexicalization of words with /r/ and /l/.

### Production by Japanese Learners.

Numerous L2 studies have indicated that learners whose L1 has no contrast between /r/ and /l/ sounds will have difficulty distinguishing them in English. The difficulty of Japanese learners in producing /r/ and /l/ has been well documented (e.g., Goto, 1971). The single liquid in Japanese, usually designated /l/, is sometimes realized as a voiced tip-alveolar flap resembling somewhat the flapped /d/ in American English “ladder.” Japanese /r/ may also be retroflexed before high front vowels. It is not permitted in consonant clusters, as is its English counterpart. Thus the Japanese learner of English must learn to produce a contrast between /r/ and /l/ according to English phonetic norms, and to produce both liquids in unaccustomed phonetic environments.

Acoustic and perceptual evidence (Hay & Fukuzawa, 1979) indicated that at least some Japanese learners of English do not produce a clear distinction between /r/ and /l/. Spectrographic analysis of /r/ and /l/ produced in English CV syllables revealed important differences between native English speakers and two Japanese learners who had lived for 3 months and 1.5 years in the United States. The Japanese subjects produced /l/ with a higher starting F3 frequency than /r/, but their /l/ versus /r/ difference was much smaller (ca. 1000 Hz) than that of English native speakers (ca. 2000 Hz, see Dalston, 1975). As a result, the /r/ produced by the Japanese learners sometimes had values appropriate for /l/. The duration of the F3 steady state portion of /r/ and /l/ produced by the Japanese talkers was about equal (103 and 108 ms), but the steady-state in native-produced /l/ is typically longer than that of /r/ (about 45 versus 30 ms, see Dalston, 1975). Finally, the Japanese learners of English produced a smaller difference in F3 transition duration between /r/ and /l/ (81 ms versus 66 ms) than native English speakers (71 ms versus 31 ms).

These acoustic measurements suggested that listeners would tend to hear both the /r/ and /l/ produced by the Japanese talkers as /l/. This was confirmed in a listening test performed by Hay and Fukuzawa (1979), which revealed that the /r/ tokens produced by the Japanese subjects was heard as /l/ more often than their /l/ tokens were heard as /r/.

Some Japanese speakers of English learn to produce English liquids adequately. Mochizuki (1981) reported that American listeners identified correctly the /r/ and /l/ sounds produced by an advanced Japanese learner of English 96% of the time. There was nearly perfect identification in all contexts that were examined (singletons in word-initial, word-final, and intervocalic positions, and word-initial clusters). However, my tabulation of the phonetic transcription data reported by Dickerson (1974) indicated that Japanese learners who had lived less than one year in the United States produced /r/ (but not /l/) more authentically in word-initial clusters than as word-initial singletons. In
an examination of the English spoken by three beginning-level Japanese learners, Dissoway-Huff (1981) found that /r/ and /l/ were judged "acceptable" (probably what we have been calling "authentic") only 79% of the time in word-initial position, compared to 67% of the time in intervocalic and word-final positions.

Perception of Japanese Learners.

Speech perception studies suggest that Japanese learners can learn to perceive English /r/ and /l/ correctly. Miyawaki et al. (1975) found that Japanese university students who had never lived in an English-speaking environment were not able, as were native speakers of English, to discriminate synthetic /r/ and /l/ stimuli at above-chance levels. However, more recent studies have shown that Japanese learners with greater English L2 experience succeed better in perceiving English /r/ and /l/. Mochizuki (1981) reported that Japanese subjects were less consistent than native English speakers in their identification of stimuli in a synthetic /ra/-to-/la/ continuum in which F3 onset was varied systematically. The Americans identified correctly the endpoint stimuli 99% of the time, compared to only 84% by Japanese subjects who had lived 6 months to 4 years in the United States. The principle error of the Japanese subjects was to label as /l/ stimuli heard by English speakers as /r/.

Mochizuki (1981) also presented native English-produced minimal pairs with /r/ and /l/ to native Japanese-speaking listeners. Correct identification averaged only 82% overall, compared to the near-perfect identification by English-speaking listeners. However, percent correct identification was higher for sounds in word-final (98%) than word-initial position (88%) which, in turn, was higher than intervocalic (73%) and word-initial clusters (69%). Sheldon and Strange (1982) obtained very similar results. They suggested that the observed pattern may have had an auditory-acoustic basis. The distinction between /r/ and /l/ may be perceived poorly in clusters because the "target" onset frequency of F3 is not achieved and the temporal differences in steady-state durations between /r/ and /l/ is reduced. It is possible that /r/ and /l/ are identified more accurately in word-final position because transition durations are longer there and the steady-state target values are more likely to be achieved.

Gillette (1980) studied the effect of a five-week classroom training program on the perception of English /r/ and /l/ by three native speakers each of Japanese and Korean. The training consisted of articulatory descriptions of /r/ and /l/, individual and group practice in producing /r/ and /l/, and auditory evaluation by the students of the extent to which their productions of /r,l/ differed from that of fellow students and the native English-speaking instructor. Their students' correct identification of /r/ and /l/ in minimal-pair words increased from 70% to 92%. The perception of /r/ and /l/ was better throughout training in final (95% correct) than in initial position (84% correct overall).

MacKain et al. (1981) found that Japanese learners with extensive experience speaking and hearing English, but not relatively inexperienced Japanese learners, identified and discriminated synthetic /r/ and /l/ stimuli much like native English speakers. The experienced Japanese learners showed a sharp crossover from /r/ to /l/ judgments as F3 onset frequency changed, and a peak in percent correct discrimination near the /r-l/ phoneme boundary.

Production and Perception of /r/ and /l/.

One question frequently raised in connection with the production of English /r/ and /l/ is whether speech perception abilities can be said to limit the learner's ability to distinguish a sound's production. If so, we would expect Japanese learners to produce /r/ and /l/ more accurately as singletons than in clusters, and in word-final than word-initial (or pre-stressed) position. However, these expectations have not been confirmed by the relatively little production data now available.

One possibility is that Japanese learners' ability to produce /r/ and /l/ correctly in word-initial consonant clusters has been overestimated. Studies (e.g., Curtis & Hardy, 1959) have shown that children learning English as an L1 appear to produce /r/ better in initial clusters than as singletons in word-initial position. Kent (1982) noted that this may stem from differences in perception rather than articulation. The production of /r/ may only appear to be facilitated in contexts such as /dr/, /gr/ and /tr/ because the release of a preceding stop "determines starting frequencies of the 2nd and 3rd formants that are compatible with the acoustic features of /r/.

A number of studies have compared the production and perception of groups of Japanese L2 learners. Cochrane (1977) examined the speech of Japanese adults and 54 Japanese children 1 to 12 years of age upon their arrival in the United States. The subjects' imitation of the speech samples (words and phrases containing /r/ and /l/ in word-initial and intervocalic position) was perceptually evaluated by native English-speaking listeners using a rating scale. The adults evidenced more omissions and substitutions, including use of Japanese /t/, than the children. The listeners accorded higher mean ratings to the English spoken by Japanese children than adults. More than half of the Japanese children were rated "unquestionably native," whereas none of the 24
Japanese adults received this rating. It is important to note, however, that even the Japanese children received lower mean ratings than native English speakers. The conclusion that the adults seemed to produce /r/ and /l/ less well than the children should be tempered by the observation that it is not certain that listeners apply the same standards in judging the speech of children and adults.

Cochrane's results suggest indirectly that the adults may have been better able to "perceive" than to "produce" /r/ and /l/. The Japanese subjects made more errors discriminating English /r/ and /l/ than English control subjects. There was not a difference in perception between children and adults, as there was in production. However, this conclusion should be accepted only cautiously because perception was assessed differently for adults and children. Whereas the adult subjects indicated which of three naturally produced stimuli (e.g., "ri", "riri," "li") was different, the discrimination of children was assessed by their imitation of the same stimuli.

Goto (1971) provided evidence that the ability of Japanese learners to produce English /r/ and /l/ is related to their ability to perceive these sounds in English words. One subset of subjects produced English /r/ and /l/ relatively well. American listeners identified 94% of their /r/ and /l/ sounds correctly. However, another subset of subjects produced these sounds poorly: only 55% of their productions of /r/ and /l/ were identified as intended. The difference in /r-l/ productive ability between the two subsets of Japanese subjects was confirmed further by the fact that even Japanese listeners identified more sounds produced by the "good" (74%) than the "poor" Japanese speakers of English (58%). This seems to rule out the possibility that the "poor" speakers were producing a difference between /r/ and /l/ that was insufficient to cue the contrast for native English speakers.

Goto (1971) also assessed perception. The Japanese subjects with "good" production discriminated /r/ and /l/ correctly more often (77%) than those with "poor" production skills (57%), suggesting that subjects who produced /r/ and /l/ adequately did so because they perceived these sounds accurately. An alternate interpretation that can not be ruled out is that the perception of some subjects was limited by their ability to correctly produce /r/ and /l/. Sheldon and Strange (1982) have pointed out that, for the Japanese subjects taken as a whole, there was a significant correlation between adequacy of production and accuracy of perception. However, the finding that has been of most interest is Goto's observation that some of the subjects with "good" production failed to discriminate /r/ and /l/ accurately.

Sheldon and Strange (1982) examined the production and perception of /r/ and /l/ by native speakers of English and Japanese. English-speaking listeners identified almost all of the /r/ and /l/ sounds produced by five Japanese learners in consonant clusters as well as those in the initial, medial, and final position of minimal pair words. These five Japanese subjects succeeded much better in identifying /r/ and /l/ (88% correct) than most of Goto's subjects, perhaps because they had lived in an English-speaking environment. The native English listeners identified only 65% of the sounds produced by a sixth Japanese subject, who was correspondingly less accurate in identifying /r/ and /l/ (74% correct) than the Japanese subjects with more adequate pronunciation.

Like Goto (1971), Sheldon and Strange (1982) found that Japanese learners may be somewhat more successful in producing a perceptually effective contrast between /r/ and /l/ than in perceiving the distinction between these sounds. This stands in contrast to the finding of English L1 acquisition research (e.g., Strange & Broen, 1980) that many children can discriminate /r/ and /l/ before they distinguish these sounds reliably in production. Sheldon and Strange suggested that their finding may be an artifact of formal training. Because of their greater cognitive sophistication, the adult Japanese subjects may have been able to develop a strategy for producing /r/ and /l/ in formal speaking contexts that is not available to young children. This suggestion is plausible in the light of the finding by Catford and Pisoni (1970) that subjects who received articulatory training for "exotic" L2 sounds produced these sounds better than subjects receiving only aural training (see also Tervoort, 1979).

We cannot reach any firm conclusions concerning the relationship between production and perception of English /r/ and /l/ by individuals learning English as an L2 based on the evidence now available. It would be especially unwise to draw any conclusion concerning causality (e.g., perception limits production rather than the reverse) based on these findings. However, it seems reasonable to conclude that both production and perception improve with experience. Another conclusion is that the production and perception of a new L2 sound or contrast may not be equivalent, at least not as measured in the studies reviewed.

This suggests that information concerning the sensory attributes of a sound and its articulatory realization may be represented independently and nonisomorphically in central phonetic representations. This inference is consistent with the view of some researchers (e.g., Carman et al., 1973; Albert & Obler, 1978) that speech production and perception abilities develop at different rates, and therefore represent different aspects of "competence." It is also consistent with Labov's (1972) finding of "partial mergers" among speakers of English dialects. For example, Costa and Mattingly (1981) found that speakers of an /r/-less American English dialect made vowels 30-40 ms longer in
“card” than “cod.” Despite this difference in production, they were unable to distinguish “card” from “cod” in recordings of their own speech or that of other native speakers of the same /r/-less dialect.

Further research is needed, especially longitudinal studies which examine both the production of /r/ and /l/ in many words and contexts, and the identification of /r/ and /l/ in the same words and contexts.

**WORD STRESS**

One of the greatest impediments to comprehension of English words is misplaced stress. Since both the placement of word stress and its implementation may differ cross-linguistically (see Berinstein, 1979), the non-native learner of English must often learn which syllable in a word is meant to be stressed, and learn to phonetically implement word stress in a new way. In English, unlike other languages such as French or Spanish, there is a fairly consistent alternation between stressed and unstressed syllables in fluent speech. English vowels in unstressed syllables tend to be relatively short in duration and reduced in quality compared to the same vowels in stressed syllables. L2 learners whose L1 makes more consistent use of temporal and spectral differences to signal stress than English, or relies to a greater extent on Fo than temporal/spectral differences, may have difficulty producing perceptually effective unstressed vowels in English.

Several studies have shown that non-native speakers more closely resemble native English speakers in their production of stressed than unstressed syllables. Adams (1979; Adams & Munro, 1978) found that in producing the same English speech materials, eight native speakers of “Asian” languages produced 50% more syllables perceived to be stressed than English native speakers. The non-native speakers were always observed to stress syllables that were stressed by native English speakers. But they also stressed words not normally stressed in English such as conjunctions and prepositions, failed to reduce unstressed syllables (e.g., the second syllable of the word “market”), and misplaced stress (e.g., placed stress on the second syllable of “faster”). Well over half the vowel pronunciation errors in a large corpus reported by Hammond (1982) were related to stress. Spanish learners frequently substituted Spanish vowels (e.g., /i/, /a/, /o/) for the schwa-like vowels in the unstressed syllables of English words. The Spanish substitute vowels seemed to depend directly on orthography. For example, “apply” was realized as /aplaI/ rather than /aplaI/.

Wenk (1983) observed that the “nativeness” of unstressed vowels in English words produced by French learners depended on the extent to which they were reduced in quality. Listeners judged vowels to be correct 99% of the time when native French speakers imitated words produced by native English speakers. This finding seems to have demonstrated that the non-native subjects were capable of producing English vowels with appropriate quality. However, vowels produced in a paragraph reading task were judged to be correct in only about 50% of instances. Part of this decrement was due to Fo differences between native and non-native speakers. Subjects seemed to produce the second (unstressed) syllable in words like “wiser” with higher Fo than the first (stressed) syllable in words like “about,” probably as the result of maintaining characteristics of French prosody in English. According to Wenk, they might also have produced unstressed syllables found at the end of a rhythmic unit with greater articulatory effort and precision than vowels found at the beginning of the unit.

Fokes, Bond, & Steinberg (1984) examined the production of English word stress by native speakers of several different languages in sets of words in which stress remained constant when affixes were added (e.g., “confess,” “confession”), or moved as the result of affixation (e.g., “confirm,” “confirmation”). Like native English speakers, all but one non-native speaker produced stressed syllables with a higher peak Fo than unstressed syllables. They also produced stressed syllables with greater amplitude than unstressed syllables, demonstrating they had accurate tacit knowledge concerning which syllables should be stressed. However, the non-native speakers conformed less closely to English phonetic norms in regards to the duration difference between stressed and unstressed syllables. Non-native speakers sometimes made stressed vowels shorter than unstressed vowels in an adjacent syllable, which is something native English speakers seldom if ever do. The non-native speakers tended to produce the unstressed second syllable in words like “confirmation” with full rather than reduced vowel quality, and only two of six non-native speakers made it shorter than the first, stressed syllable.

Hutchinson (1973) found that Spanish learners produced only a small duration difference between stressed and unstressed syllables in English words upon their arrival in the United States. The difference they produced resembled in magnitude the difference observed in Spanish. However, the L2 learners produced a much greater temporal difference between stressed and unstressed English syllables after a six-month intensive English course. This change may have helped improve their English accent, for Hutchinson noted a high positive correlation between the degree to which stressed and unstressed syllables differed in duration and global foreign accent ratings.

Another possibility, of course, is that subjects who stressed syllables in an English-like way may have also improved other aspects of sound
and syllable articulation which contributed independently to the improved accent ratings. This interpretation is consistent with results reported by Elsendoorn (1983b). He manipulated the duration of vowels in English words produced by Dutch speakers. Changing the duration of vowels so that they equalled the average duration of vowels produced by native English speakers did not result in higher acceptability judgments by English-speaking listeners.

In summary, the data now available indicates that non-native speakers often fail to observe one very important rhythmic property of English, namely the reduction of unstressed vowels. The general tendency for non-native speakers to produce unstressed syllables with prominence in speaking English may be due to uncertainty concerning lexical stress or the rhythm of English. For example, Bond and Fokes (1985) recently noted that non-native speakers differ from native English speakers in reducing vowel duration in a word as suffixes are appended to it (e.g., as in speed, speedy, speedier). The findings pertaining to the implementation of stress may be related to the general tendency for non-native speakers to speak with a slower rate of articulation (see, e.g., Flege, 1979; Niemi, 1984). In addition to decrements in intelligibility, failure to unstress vowels may lead to a “staccato” rhythm. Much more research is needed to clarify the effect on intelligibility and degree of accent of divergences from English phonetic norms for stress implementation.

VOWEL TIMING

The controlled variation of vowel duration may serve various phonetic functions in different languages. It may contribute importantly to the distinction between phonemically long (/V:/) and short (/V/) vowels, between tense and lax vowels (e.g., /i/ versus /I/ in English), or between stressed and unstressed vowels. It may serve to indicate word and phrase boundaries and, at least in English, help cue the voicing feature of a following obstruent. In this section we will review studies pertaining to the control of vowel duration in L2 speech production, focusing on instances where L1 and L2 differ in vowel timing. We will also review several studies examining the perceptual effect of vowel duration.

Intrinsic Timing.

Speech timing studies have shown that many factors influence predictably the basic or “inherent” duration of vowels (Lehiste, 1970). Some are common to many languages. For example, it seems to be true that “low” vowels like /a/ are universally longer than “high” vowel like /u/. Another seeming universal is that vowels are longer preceding fricatives than stops, and preceding dental/alveolar or velar than labial stops. One possible explanation for the widespread occurrence of such vowel timing effects is that they follow intrinsically from the gestures used to produce vowels rather than from control parameters established during L1 acquisition. If so, we should never observe differences between native and non-native speakers regarding intrinsic timing factors because they do not need to be learned. This expectation has been supported by several instrumental studies of L2 speech production (e.g., Suomi, 1976; Flege, 1979; Elsendoorn, 1980, 1982; Mitleb, 1981).

It would be useful to extend these studies to the production of new L2 vowels. If certain temporal properties of vowels are intrinsic to their mode of articulation, they should emerge rapidly in L2 speech. So, for instance, a Chinese or Dutch learner of English should produce English /æ/ with longer duration than English /i/ even in the earliest stages of L2 learning, despite the fact that /æ/ is a new vowel which does not occur in Chinese or Dutch. Similarly, although Chinese words do not end in /s/ or /t/, Chinese learners of English should make vowels longer before /s/ than /t/ in English words just as soon as they succeed in producing these obstruents in word-final position.

Extrinsic Timing.

Other characteristics of vowel timing probably need to be learned. One straightforward example of an “extrinsic” timing factor is the phonemic length contrast used to distinguish vowel phonemes in about half the world’s languages. Phonemically “long” vowels are generally twice as long as corresponding short vowel phonemes.

The extrinsic timing factor which has received the most attention in L2 studies is English differential vowel duration. English pre-pausal vowels are typically 40% (5–70 ms) longer before voiced (e.g., /b,d,g/) than voiceless (e.g., /p,t,k/) obstruents. These large temporal differences do not appear to be a consequence of the laryngeal gestures used to produce voiced versus voiceless obstruents. The magnitude of the effect seems to be larger in English than any other language examined to date, and serves a perceptual function. Many studies (e.g. Krause, 1982) have shown that native English children and adults tend to judge an obstruent to be voiced when it is preceded by a relatively long vowel, especially in noisy listening conditions which obscure spectral cues to the voiced-voiceless contrast. Vowel duration does not seem to affect voicing judgments for speakers of languages like Dutch in which differential vowel duration is absent or much smaller in magnitude than it is in English (Slis & Cohen, 1969; Staatsen & Leijten, 1976).
Phoneticians (e.g., Chen, 1970) have concluded that vowel duration has come to serve as a perceptual cue to the voicing feature of a following obstruent because a small intrinsic timing factor was exaggerated in English. If so, it is likely that English differential vowel duration is learned and represented centrally. For native speakers of English, such information might be stored in the central phonetic representations for consonants which, in turn, could be manifested in production of the preceding vowel as phonetic strings are implemented.

Many studies have demonstrated that listeners possess great sensitivity to the temporal properties of vowels in their L1. For example, Abramson (1962) found that Thai native speakers showed a sharp crossover from phonemically short to long vowel judgments as the duration of vowels in the stimuli was increased in small increments. Jonasson and McAllister (1972) showed that Swedish subjects were able to correctly reject as "foreign" Swedish words that had been synthesized with vowel and consonant durations that were not typical of spoken Swedish (see also Johansson, 1975). Nooteboom (1973) had Dutch subjects adjust the vowel duration in auditorily presented words until they sounded "correct." The Dutch subjects showed great consistency (SD's of 4–6 ms for 20 trials), making long vowels about 100 ms longer than short vowels. They even varied the vowel duration in words of different lengths to match the pattern of temporal variation seen in speech production (vowel duration decreased when the total word duration was increased).

Another extrinsic timing factor that has received attention in L2 studies is the temporal difference between English tense (/i/, /u/, /e/) and lax (/I/, /u/, /e/) vowels. Cross-language comparisons indicate that a tense vowel is not always longer than the corresponding lax vowel. One study (Elsendoorn, 1980) showed that English /i/ was 82% (108 ms) longer than /I/, whereas Dutch /i/ was only 9% (10 ms) longer than Dutch /I/. A number of studies have shown that duration serves as a secondary perceptual cue to the identity of tense versus lax vowels (e.g., Bennett, 1968; Ainsworth, 1972). Thus, as for English differential vowel duration, the duration difference between tense and lax vowels in English is likely to be a property that is learned and stored centrally. Assuming that the same temporal contrast does not exist in L1, we can expect to see differences between native and non-native speakers of English in regards to the temporal distinction between tense and lax vowels.

**English Differential Vowel Duration.**

Several studies with similar methodologies have examined vowel duration as a function of the voicing characteristic of the following stop consonant. In these studies, native and non-native speakers have read minimal pair words from lists, either in isolation or embedded in a fixed carrier phrase. The absolute durations reported for vowels varied across studies, probably as the result of different phonetic contexts, speaking rates, degrees of stress, and so on. The results have nonetheless been consistent in showing that native English speakers produce a larger magnitude of differential vowel duration than non-native speakers.

Several studies examined the English spoken by native speakers of Arabic and French. In Arabic, there is no difference in the duration of vowels preceding voiced versus voiceless stops (Flege, 1979), whereas French seems to possess a significant yet somewhat smaller effect than English (Mack, 1982; see also Flege & Hillenbrand, 1985). To produce English vowels authentically, native speakers of Arabic must learn to produce differential vowel duration, and French speakers must learn to increase the magnitude of the effect found in their native language.

It may be easier for L2 learners to establish control of a new timing parameter than to modify an L1 timing parameter. However, this hypothesis is not supported by the evidence which now exists with regard to English differential vowel duration. Flege and Port (1981) found that both inexperienced and experienced Saudi Arabians (9 versus 38 months in the United States) produced a smaller magnitude of differential vowel duration (5% or 6 ms versus 6% or 8 ms) than native English speakers (20% or 32 ms). The small effects observed in English for these L2 learners closely matched those observed in their production of Arabic. That is, they showed little evidence of having modified their timing of vowels in L2. Very similar results have been reported by Port and Mitleb (1983) for Jordanians who had never lived in an English-speaking country, and by Mitleb (1981) for Jordanians who had lived for at least two years in the United States.

A recent study by Fokes, Bond, and Steinberg (1985) examined vowel duration in the English spoken by 12 native speakers of Arabic aged 2.5 to 11 years (M=5.9 years) who had lived for 2 months to 4 years in the United States. The children made vowels significantly longer before /p/ and /t/ than before /b/ and /d/ in English words. This might mean that children learn English differential vowel duration more readily than adults, but Krause noted a much larger effect for native English six-year-olds than the effect Fokes et al. (1985) observed for the Arabic youngsters (approximately 100 ms versus 60 ms).

Mack (1982) measured vowel duration in words spoken in French or English by monolinguals, and in French and English by native speakers of French who had lived in the United States for 8 months to 4 years. As expected, the English monolinguals showed a considerably
larger voicing effect than the French monolinguals (85% or 131 ms versus 37% or 49 ms). Like the adult Arabic L2 learners mentioned earlier, the French L2 learners showed little evidence of having approximated the phonetic norms of English. The magnitude of the voicing effect they produced (27% or 48 ms) resembled the one seen for French monolinguals much more than the one seen for English monolinguals.

One reason why neither the Arabic nor French L2 learners seem to have approximated the phonetic norm of English is that pre-established patterns of L2 vowel timing carried over into their production of English L2. Both Arabic and French have words ending in voiced and voiceless obstruents. Perhaps the subjects of both L1 backgrounds maintained a previously established (L1) timing pattern in L2 production.

The same thing could not be said with regard to Dutch and Finnish learners of English, for neither Dutch nor Finnish possess a contrast between voiced and voiceless obstruents in word-final position. However, Dutch and Finnish learners do not seem to learn English differential vowel duration. Elsendoorn (1980, 1982) found that native English speakers produced a substantially larger temporal distinction (87% or 122 ms) in vowels preceding voiced versus voiceless obstruents in English words than native speakers of Dutch (8% or 13 ms). Suomi (1976) reported that Finns majoring in English at the university produced a very small magnitude of differential vowel duration (16% or 25 ms) that scarcely exceeded that of Finns who had not studied English beyond high school (11% or 9 ms). Although the voiced-voiceless difference produced by the university students reached statistical significance (as for some of the other groups mentioned earlier), it was considerably smaller than that observed for native English speakers (73% or 99 ms).

The seven studies just reviewed indicate that non-native speakers do not seem to learn English differential vowel duration. One other study did, however, show similarity between native and non-native speakers of English. Mitleb (1981) examined the English vowels spoken by Germans. Like Dutch, German is said to have a rule which devoices voiced obstruents (e.g., /b,d,g/) in word-final position. Mitleb's subjects were newly arrived graduate students at Indiana University who had lived an average of only 5 months in the United States at the time of the study. They made vowels 31% (55 ms) longer before voiced than voiceless stops in one-syllable words, compared to 33% (53 ms) for native English speakers. The German subjects produced a much larger effect in English than German words (15% or 23 ms), suggesting they modified vowel timing in L2.

The German's success could conceivably have been due to relatively great motivation to pronounce English authentically, or to better English L2 instruction. It probably cannot be explained as resulting from more experience with English. As mentioned earlier, Arabs who had lived much longer in the United States than the German subjects did not produce differential vowel duration (Flege & Port, 1981; Port and Mitleb, 1983). Mitleb (1981) suggested that the Germans did not actually need to learn the large-magnitude effect characteristic of English, for a vowel duration effect comparable to the one in English occurs in German two-syllable words. The German L2 learners may simply have generalized differential vowel duration to a new phonetic context (i.e., one-syllable words) in English.

Another point of agreement among the vowel timing studies just mentioned is that native and non-native show more similarity for vowels preceding /p,t,k/ than /b,d,g/. Linguists have proposed that English possesses a phonological rule which lengthens vowels before /b,d,g/. The empirical evidence is consistent with this view. The small magnitude of differential vowel duration might therefore be regarded as stemming from the failure of L2 learners to learn a phonological rule of English. It is also possible, of course, that they simply failed to learn a parameter associated with the production of final obstruents (one which is stored in the central representations for stops, but affects the duration of adjacent vowels in speech production).

Another possible explanation is that L2 learners do not produce English differential vowel duration because they do not perceive it. We suggested earlier that vowel duration may not act as a perceptual cue to the voicing feature of final obstruents for individuals whose L1 does not possess a large magnitude of differential vowel duration. However, this hypothesis was largely disconfirmed by Elsendoorn (1981), who systematically shortened the vowels in naturally produced CVC English words ending in /b,d,g,z,v/. Elsendoorn's natural-edited stimuli were presented in a forced-choice task to native speakers of English and to six groups of Dutch subjects differing in age and amount of English-language experience. The dependent measure was perceptual "uncertainty," that is, the number of stimuli that were not judged consistently as ending in a voiced or voiceless obstruent. Most (but not all) native English subjects showed a sharp crossover from voiced to voiceless stop judgments as vowel duration decreased. Experienced Dutch speakers of English (university students and staff) resembled the native English speakers, suggesting that differential vowel duration can acquire perceptual cue value for L2 learners. Vowel duration appeared to have little effect on the stop voicing judgments of less experienced subjects (Dutch high school students). Like a small minority of the native
English speakers, they seemed to pay primary attention to the quality of the preceding vowel in making their voicing judgments. However, in another study Elsendoorn (1984) found that even inexperienced Dutch speakers of English may have tacit knowledge that English vowels are longer before /b,d,g/ than /p,t,k/. He had Dutch high school students aged 15 to 18 years adjust the duration of the vowel in auditorily presented CVC English words until they seemed to be “correct.” Like native English-speaking subjects, the Dutch subjects’ inner criteria led them to adjust vowels to be considerably longer in words ending in /b,d,g/ than /p,t,k/. This suggested that the Dutch subjects were aware of English differential vowel duration, even if they did not produce it in speaking English. It also demonstrated the truism that assessing the perceptual skills in different tasks may tap different processing levels which, in turn, yields different results.

**English Vowel Tensity.**

Several studies examined the duration of tense and lax vowels in English words spoken by non-native speakers. Mitleb (1981) found that native speakers of Arabic made tense vowels (/i, u, e/) 47% longer than lax vowels (/æ, ə/) in CVC English words. The magnitude of their vowel tensity effect was considerably greater than that of native English speakers (15% or 24 ms). Suomi (1976) examined vowels produced by native English speakers and Finns differing in English-language experience (i.e., Finns who studied English only in high school or had majored in English at the university). Both groups of Finns made tense vowels considerably longer (about 106% or 110 ms) than lax vowels. As for the Arab learners just mentioned, the magnitude of their vowel tensity effect was considerably greater than that of native English speakers (58% or 68 ms).

It was suggested earlier that L2 learners may reinterpret the nature of the distinction between L2 sounds. Neither Finnish nor Arabic has a phonemic contrast between pairs of tense and lax vowels. However, both languages contrast phonemically long and short vowels (e.g., /i/ versus /i:/) by means of a substantial duration difference. The spectral difference between Finnish /i/ and /i:/ closely resembles that distinguishing English /i/ and /i:/ (Wiik, 1965). The Finns’ 100-ms difference between English /i/ and /i:/ corresponded closely to the 100-ms difference distinguishing long versus short vowel phonemes in Finnish (Wiik, 1965). Similarly, Mitleb (1981) noted that Arabs’ temporal contrast between English /i, ɪ/ closely matched the temporal contrast they produced between Arabic /i/ and /i:/ these considerations suggest that the Finns and Arabs reinterpreted the English tense/lax distinction as a phonemic length distinction. If so, they did not learn a new temporal distinction or modify an L1 distinction, but simply maintained a pre-established timing pattern in L2.

This interpretation may also apply to Elsendoorn’s (1980) results for Dutch native speakers who produced /i/ and /i:/ in isolated English and Dutch CVC words. The Dutch native speakers made /i/ considerably longer than /i:/ (55% or 72 ms), although they produced little difference between Dutch /i-/ (8% or 10 ms). The tense/lax difference produced by the Dutch speakers in English words is somewhat smaller than the temporal difference we have come to expect for phonemic length differences, and it was also smaller than the tense/lax difference produced by native speakers of British English (82% or 108 ms).

This finding seems to suggest that the Dutch learners partially approximated the English temporal distinction between tense and lax vowels. Other evidence suggests, however, that the Dutch learners may have re-interpreted the English /i/ versus /i:/ distinction as a phonemic length opposition. In Dutch there are four pairs of long-short vowel phonemes which differ both in duration and spectral quality. The short Dutch /i/ enters into a phonemic length opposition with /e/-. Dutch /i/ does not enter into a phonemic length opposition, but its duration is comparable to that of other short vowel phonemes of Dutch. The percentage by which /i/ duration exceeded that of /i:/ in the English spoken by Dutch learners in Elsendoorn’s (1980) study closely resembled the difference between Dutch /e:/ and /i:/ (53%) reported by Nooteboom & Slis (1972).

**Phonemic Length Distinctions.**

For native speakers of non-quantity languages like English, the distinction between phonemically long and short vowels represents a new distinction that must be learned. There is little evidence which bears directly on the question of whether this dimension can be learned in an L2. One study (Jonasson & McAllister, 1972) examined production of a phonemic length distinction in Swedish by a native speaker of English. The subject spoke Swedish with a “typical American accent.” Not surprisingly, he did not produce a clear temporal distinction between long and short Swedish vowels, often making phonemically long vowels shorter than phonemically short vowels. My own analysis of the data suggests that the American subject produced Swedish vowels with the duration appropriate to the most spectrally similar English vowel, even though the Swedish long-short opposition is clearly marked in orthography. This suggests the substitution of English for Swedish vowels as the result of interlingual identification.
CONSONANT TIMING

In English the closure duration of voiceless obstruents (e.g., /s/ or /p/) is longer than that of homorganic voiced obstruents (/z/ or /b/) in word-final position. This temporal contrast is not universally present in human languages (see Flege & Brown, 1982), so it probably needs to be learned. Unlike the results of studies examining vowel timing, four existing studies agree in showing that this temporal contrast is learned in L2.

Two studies examined the English spoken by Arabs. Flege and Port (1981) measured the duration of /p,t,k/ and /b,d,g/ in the final position of one-syllable Arabic and English words. Native Arabic speakers produced no closure duration difference between voiced and voiceless stops in Arabic words. Inexperienced Arab learners (eight months in the United States) did not produce a significant difference between /p,t,k/ and /b,d,g/ in English words. Experienced Arab learners (38 months in the United States), however, did produce a temporal distinction between /p,t,k/ and /b,d,g/, albeit a smaller one (16% or 10 ms) than a native English control group (36% or 19 ms). Similarly, Port and Mitleb (1983) found that Jordanians who had lived for 12 to 16 months in the United States made /p/ longer than /b/ (by 56% or 33 ms), while a matched group of Jordanians who had never lived in an English-speaking environment did not (producing only a 5%, or 6 ms, difference).

Suomi (1976) measured the duration of /p,t,k/ and /b,d,g/ in the final position of CVC English words. (No comparable measure could be made for Finnish words since Finnish does not possess voiced stops in word-final position.) Finns who had studied English only in high school produced a small and nonsignificant contrast between /b,d,g/ and /p,t,k/ (9% or 9 ms), while those who majored in English at the university made /p,t,k/ significantly longer (by 29% or 26 ms) than /b,d,g/. In fact, their contrast was similar to the one produced by native English speakers (36% or 27 ms).

Mitleb (1981) found that German learners made /p,t,k/ significantly longer than /b,d,g/ in both two-syllable (29% or 19 ms) and one-syllable English words (17% or 14 ms). However, this finding represents less certain evidence of phonetic learning than the studies just mentioned. Although these subjects produced only a nonsignificant 8 ms (11%) difference at the end of one-syllable German words, they made /p,t,k/ longer (by 56% or 29 ms) than /b,d,g/ in two-syllable German words. Thus they may simply have generalized a temporal phonetic distinction already present in L1 rather than learning a new distinction in L2.

We saw earlier that L2 learners may use L1 vowel timing distinctions in L2 even if it is inappropriate to do so in L2. The same seems to be true for consonant timing. Flege and Port (1981) measured the closure duration of /p,t,k/ and /b,d,g/ in the word-initial (unstressed) position of comparable Arabic and English words. Native English speakers produced no difference between word-initial /b,d,g/ and /p,t,k/ in English words. However, native speakers of Arabic made /p,t,k/ significantly longer than /b,d,g/ in the initial position of English words, just as they did in Arabic words. Thus the Arabs maintained an Arabic temporal distinction in English.

I know of no empirical evidence concerning the ability of native speakers of English to learn the temporal contrast between /d,g/ and /t,k/ in the initial position of Arabic words, nor is there much evidence pertaining to the perception of consonant timing by non-native speakers. Johansson (1975) found that Swedish listeners’ acceptability judgments of words ending in stop consonants depended on the duration of stop closure. The same was not true for English-speaking listeners. It would be interesting to determine whether native English speakers who learn Swedish become more sensitive to the duration of final stops.

One study suggested that L2 learners may not use the duration of a final obstruent as a perceptual cue to obstruent voicing. Flege & Hillenbrand (1985) examined the identification of the members of a “peas” to “piece” continuum in which vowel and final fricative duration were each varied in five 50-ms steps. In English and French, vowels are considerably longer before word-final /z/ than /s/ and /z/, at the same time, the frication noise of /s/ is considerably longer than that of /z/. Native speakers of English and French showed large increases in /z/ judgments as vowel duration increased, and smaller increases in /z/ judgments as fricative duration decreased. The labeling responses of native speakers of two languages in which there is no contrast between /s/ and /z/ (Swedish and Finnish) were also gathered. Swedes and Finns showed a large effect of vowel duration, but not even those who had lived for a year in the United States showed an effect of fricative duration. This suggested that it might be difficult for L2 learners to integrate multiple acoustic cues for a new phonetic contrast in L2, especially when a cue which is as salient as vowel duration can be used to divide the stimulus continuum into two categories.

Additional studies of segmental timing are needed to establish the extent to which learners become sensitive to temporal parameters used in L2 but not L1. An important issue is whether it is easier to abandon or modify L1 timing patterns than to learn new ones in L2. Future research should try to determine why L2 learners seem more likely to
I serve to distinguish it from members of other categories (e.g., learn a new temporal specification for word-final stops than stressed vowels.

VOICE ONSET TIME

Many studies of L1 and L2 speech production have focused on the VOT dimension in prevocalic stop consonants. VOT is a temporal measure of the relative onset of glottal and supraglottal articulatory events. According to convention, it is measured as the time, in milliseconds, from the release of stop constriction and the beginning of periodic vibration of the vocal folds (see also Winitz, LaRiviere, & Herriman, 1975). Cross-language research (e.g., Lisker & Abramson, 1964) has shown that pairs of homorganic stops in most (if not all) languages differ in terms of VOT. The VOT values associated with a single category (e.g., /p/) are usually produced with “short-lag” VOT values. In short-lag stops, which are assigned small positive VOT values, phonation begins soon after stop release. In languages like Chinese and English the /b,d,g/ categories are realized with short-lag VOT values similar to those measured for /p,t,k/ in languages like Spanish and French. Chinese and English /p,t,k/ are realized as “long-lag” stops. They have significantly longer (positive) VOT values than the /p,t,k/ in languages like Spanish and French.

Despite the tendency of stops to fall into one of three “modal” ranges (i.e., lead, short-lag, or long-lag), the actual mean VOT value measured for stops in any one range may also vary somewhat across languages. For example, the “long-lag” /t/ and /k/ of Arabic have somewhat shorter VOT values than corresponding voiceless stops in English which, in turn, have somewhat shorter values than the /t/ and /k/ of Danish (Flege, 1979; Christensen, 1984). This means that individuals must learn language-specific VOT values to ensure authenticity and adequacy of production.

The number of phonetic categories perceived by monolinguals in a continuum of stops differing in VOT depends on the phonology of their native language. For example, Lisker and Abramson (1970; Abramson & Lisker, 1970; 1973) examined labeling of synthetic CV stimuli by native speakers of English, Spanish, and Thai. The Thai subjects divided three continua (/ba-pa/, /da-ta/, /ga-ka/) into the three stop categories found in Thai (i.e., lead, short-lag, long-lag), whereas the

English and Spanish subjects used only two labels corresponding to the two phonetic categories of their L1.

For languages with only two phonetic categories differentiated by VOT, the point along the continuum where judgments shift from predominantly voiced to voiceless, known as the phoneme boundary, may differ according to the phonetic norms of the language. For example, Christensen (1984) examined labeling by native speakers of Danish and English, two languages in which /p/ is realized as a long-lag stop. The phoneme boundary occurred at significantly longer VOT values for Danish than English subjects, probably because Danish /p/ is realized with somewhat longer VOT values (about 100 ms) than English /p/ (about 70–90 ms in isolated words).

The perceptual effect of VOT closely mirrors differences observed in the production of voiced and voiceless stops, even at a very early age. For example, Lisker and Abramson (1970) found that the /b/-/p/ phoneme boundary of monolingual native English speakers occurred at about 25 ms. These subjects produced /b/ with VOT values shorter than 25 ms, and /p/ with VOT values greater than 25 ms. Data reported by Williams (1977a) indicated a close, but different, match between production and perception for monolingual native Spanish speakers. The /b/-/p/ phoneme boundary for several groups of Latin Americans was about −7 ms. Correspondingly, these subjects produced Spanish /b/ with lead VOT values shorter than −7 ms, and /p/ with short-lag VOT values that were longer than −7 ms.

One notable mismatch between production and perception is evident in data reported by Lisker and Abramson (1970; Abramson & Lisker, 1973) for native speakers of Spanish. The /b/-/d/, /t-d/, and /k/-/g/ phoneme boundaries of the native Spanish-speaking subjects occurred at much greater values (about 20 ms) than that observed for monolingual Spanish subjects in Williams’ (1977a) study (−7 ms). These Spanish subjects produced /b/ with lead VOT values and /p/ with short-lag VOT values of about 0–20 ms. This means that they realized stops like /b/ and /p/ with VOT values that probably would have led them to label both as belonging to just one category (i.e., /b/).

This mismatch is puzzling for two reasons. First, it is widely accepted (e.g., Lisker & Abramson, 1970, p. 565) that the perception of children learning L1 gradually evolves until it matches the phonetic norms for speech production of the surrounding linguistic community. The subjects examined by Lisker and Abramson (1970) were adults. Second, Williams (1977a) did find a close match between production and perception for monolingual Spanish subjects using one of the same synthetic stimulus continua employed by Lisker and Abramson (1970).
The discrepancy between the two studies may have been due to linguistic experience. The Spanish subjects examined by Lisker and Abramson (1970) were tested in the United States and all spoke English to some extent, whereas the Spanish subjects examined by Williams (1977a) were tested in their own country and presumably did not speak English.

The most important evidence supporting this interpretation of the apparent discrepancy just noted is based on speech discrimination tests. "Discrimination" refers to a listener's ability to determine whether two stimuli are physically the same or different. There is often a peak in percent correct discrimination scores for pairs of speech stimuli straddling the phoneme boundary established in an identification test. This means that two stimuli given different phonetic category labels are easier to discriminate than a pair of stimuli which differ physically in the same manner but have been given the same category label. Abramson and Lisker (1970, p. 573) concluded that the discriminability of speech stimuli is "basically determined by specific language experience" because discrimination peaks often correspond closely to language-specific phoneme boundaries.

However, L2 data suggest that factors in addition to language experience are important in determining which pairs of stimuli are discriminated accurately in AX discrimination tests. Williams' (1977a) data for monolingual Spanish subjects revealed a sharp peak in the percentage of correct discrimination near the Spanish phoneme boundary (−7 ms) and a second, less prominent peak near the English phoneme boundary (about 20 ms). Since these Spanish subjects had little or no experience with the English phonetic contrast between short-lag and long-lag stops, one would not have expected to see the second discrimination peak if linguistic experience alone shaped their perception of stops.

A two-peak pattern is also evident in the discrimination data reported by Abramson and Lisker (1973) for Spanish speakers of English. The authors present no grouped data, but inspection of the individual subject data indicates a discrimination peak near the English boundary (about 20 ms) and less prominent peaks near the phoneme boundary (−7 ms) established for monolingual Spanish speakers by Williams (1977a).

A similar pattern is evident in the speech sound identification data reported by Caramazza et al. (1973). They presented CV stimuli differing in VOT to native speakers of English and to French Canadians who either did or did not speak English. The mean percentage of voiceless stop identifications by native English subjects showed a monotonic increase as VOT increased but the identification functions of the native French speakers were nonmonotonic. For them, voiceless stop judgments increased at about −5 ms along the VOT continuum, but usually did not reach 50%. This was followed by a decrease in voiceless stop judgments and a second, more dramatic increase in voiceless judgments at about 20 ms (which did cross over the 50% "phoneme boundary" mark).

The first increase in voiceless stop judgments (at around −5 ms) was much more pronounced for monolingual than bilingual French subjects. Caramazza et al. (1973) interpreted their data to mean that VOT is not a sufficient perceptual cue to the voicing feature in stops for French monolinguals. Their conclusion was based on the fact that the identification functions did not cross 50%. (Phoneme "boundaries" have been calculated traditionally as the point along a continuum where both of two possible category labels are used 50% of the time.) However, another interpretation is that the French subjects focused attention on the acoustic distinction between voiced and voiceless stops found in both French and English.

The French monolingual subjects were likely to have been exposed to the stops of English. The fact that the shift in judgments was smaller at the French than English phoneme boundary suggested that the English phonetic distinction was perceptually more salient to the listeners. Williams (1980) noted that the acoustic parameters distinguishing short-lag from long-lag stops (the English phonetic contrast) is far richer than that distinguishing lead from short-lag stops (such as those found in Spanish and French). Recent research employing speech and nonspeech stimuli with human (e.g., Soli, 1983) and nonhuman subjects (e.g., Kuhl & Padden, 1982) corroborates this observation. They provide evidence that listeners' innate auditory capabilities probably make them relatively more sensitive to the acoustic difference distinguishing voiced from voiceless stops in English than in languages like Spanish or French.

This same conclusion is supported by infant speech perception research. Eimas (1975) found that infants being reared in an English-speaking environment discriminated pairs of stimuli straddling the (adult) English phoneme boundary better than stimuli drawn from the same (adult English) categories. Two studies addressed the issue of whether infants' performance depended on early environmental experience with speech. Lasky, Syrdal-Lasky, and Klein (1975) and Streeter (1976) tested infants in Spanish and Kikuyu-speaking environments, respectively. The infants tested were unlikely to have heard a distinction made between short- and long-lag stops resembling the phonetic distinction between voiced and voiceless stops in English. The infants performed much like the English infants studied by Eimas (1975), suggesting that linguistic experience is not needed to discriminate between English voiced and voiceless stops.
The existing evidence concerning infants' ability to discriminate lead versus short-lag stops, on the other hand, is much less certain. Infants being reared in an environment where this acoustic difference underlies a phonetic distinction can discriminate such stops accurately (Lasky et al., 1975). Three studies have examined the ability of infants being reared in an English-speaking environment, where the acoustic difference between lead and short-lag stops does not underlie the phonetic distinction between the voiced and voiceless categories. The results of a study by Eimas (1975) were equivocal; the results reported by Eilers, Gavin, and Wilson (1979) were negative. A third study, designed to provide a more sensitive measure of infants' auditory abilities (Aslin, Pisoni, Hennessy, & Perry, 1981), indicated that infants being reared in an English-speaking environment can discriminate lead from short-lag stops. However, taken together, the three studies suggested that it may be somewhat more difficult for infants to discriminate the Spanish lead versus short-lag distinction in stops than the English short-lag versus long-lag distinction.

In summary, adult L2 speech perception data suggest that native speakers of languages like Spanish and French are auditorily sensitive to the acoustic contrast between short-lag and long-lag stops in addition to being phonetically sensitive to the acoustic contrast between lead and short-lag stops. Learning English as an L2 seems to heighten their sensitivity to the acoustic contrast between short- and long-lag stops and, at the same time, to diminish their sensitivity to the distinction between lead and short-lag (LI) stops.

VOT in Child Speech.

A longitudinal study by Kewley-Port and Preston (1974) confirmed the earlier observation of Jakobson (1968) that the first stops produced by children tend to be short-lag stops, no matter what the language spoken in the surrounding environment. They examined spectrographically /t/- and /d/-like stops produced in prevocalic position by three children being reared in an English-speaking environment. The first sounds recognizable as stops appeared in babbling around the age of 5 to 6 months, with no evidence of a concentration of values at any of the three “modal” VOT ranges observed in adult speech.

The wide range of values measured (−150 to +150 ms) suggest that, at this age, infants are unable to coordinate the timing of laryngeal and supraglaryngeal gestures. Within the next several months, however, the children began to produce short-lag stops with increasing frequency, suggesting they were beginning to develop rudimentary control of laryngeal timing. Kewley-Port and Preston (1974) hypothesized that the seemingly “universal” preference for short-lag stops may occur because short-lag stops, unlike long-lag stops, do not require the vocal folds to be adducted at a precise time with respect to stop release.

Children in the Kewley-Port and Preston (1974) study began distinguishing /p,t,k/ and /b,d,g/ by means of VOT in their second year. They produced an increasing percentage of stops with long-lag VOT values; and stops with lead VOT values also began to appear more frequently. When just those syllables recognizable as English words were examined, it was found that /t/-initial words had somewhat longer VOT values than /d/-initial words. However, there was still no clear bimodal distribution of values corresponding to the voiced and voiceless stop categories seen for adult native speakers of English.

An examination of several older children by Kewley-Port and Preston (1974) suggested that by the age of four years, children distinguish /p,t,k/ and /b,d,g/ by means of VOT, although some differences between adults and children may still remain. The children’s range of values for /t/ was much greater than that of adults because their /t/ was often produced with short-lag VOT values appropriate for /d/, and they were much more likely than adults to produce /d/ with VOT values in the long-lag range appropriate for /t/. Zlatin and Koenigsknecht (1975) examined the stops produced by 2- and 6-year-old children. While even the 2-year-olds produced a difference in VOT between /p-b/, /t-d/, and /k-g/, they showed more overlap between voiced and voiceless stops than the older children. The children produced /p,t,k/ with somewhat shorter mean VOT values than adults, and with a wider range of values. This means that the children’s overlap in VOT was due almost entirely to their occasional production of /p,t,k/ with short-lag values appropriate for /b,d,g/. However, data reported for 3-year-olds by Bailey and Haggard (1980) indicated little overlap in VOT values between /p,t,k/ and /b,d,g/.

Similar tendencies were observed (Clumeck, Barton, Macken, & Huntington, 1981) in the speech of children learning Chinese, a language which closely resembles English in terms of VOT. However, Macken (1980) found that children learning English as their LI distinguished word-initial /b,d,g/ and /p,t,k/ sooner than Spanish-learning children. She observed that whereas most 2-year-olds learning English as their LI produced a statistically significant difference in VOT between /p,t,k/ and /b,d,g/, few Spanish-learning children as old as four years of age did so.

Macken speculated that the difference between children being reared in English and Spanish-speaking environments was due to phonetic input. There is often no VOT distinction between the /p,t,k/ and /b,d,g/ categories in Spanish, where /b,d,g/ are often realized as continuants...
rather than stops. It seemed that, as a result, the Spanish children were learning to distinguish /b,d,g/ from /p,t,k/ by producing a stop versus continuant distinction as opposed to a VOT distinction.

Macken (1980) offered a stage-oriented model of VOT development to describe what she had measured in child speech (see also Kewley-Port & Preston, 1974). For a time, children may produce a statistically significant VOT difference between /p,t,k/ and /b,d,g/ that is less reliable and perhaps more difficult to discriminate than the differences produced by adult monolingual speakers of English. Macken regarded the small but significant voiced-voiceless differences in stops produced by certain children as evidence that they were aware of the phonological contrast between /b,d,g/ and /p,t,k/. Children later enhance the VOT difference between /p,t,k/ and /b,d,g/ by exaggerating (i.e., overshooting) the phonetic norm for /p,t,k/. This suggests that, at some point in speech learning (probably before the age of 3 years, see Bailey & Haggard, 1980) the child suddenly discovers the perceptual importance of VOT for distinguishing phonetic categories, finds an articulatory means for implementing the perceived difference between categories, or both. The child eventually reduces VOT in voiceless stops and begins producing /p,t,k/ according to the phonetic norms of English.

There is less evidence available concerning children's use of VOT as a perceptual cue than there is production data. Zlatin and Koenigsknecht (1976) examined the labeling of synthetic continua (e.g., "bear" versus "pear") in which the VOT of the initial stop was varied. There was no significant difference between adults and children in the location of the phoneme boundaries but there were other important age-related differences. A measure of "boundary width" (i.e., the range of stimuli not predominantly labeled as either voiced or voiceless) indicated that the adults were more consistent in their labelling than the children (see also Zlatin & Koenigsknecht, 1975; Simon & Fourcin, 1978). The adult subjects' identification functions were more nearly monotonic than those of the children. The two- and six-year-olds often showed two or more 50% crossovers (18% and 8%, respectively), while the adults generally manifested a single 50% crossover from predominantly voiced to voiceless responses.

Using careful testing procedures with an even larger number of children 20 to 50 months of age, Bailey and Haggard (1980) also found that response uncertainty decreased with age. However, unlike Zlatin and Koenigsknecht (1976), they found that the phoneme boundaries occurred at increasingly longer VOT values as a function of age.

The Modification of VOT.

There are two major reasons why native and non-native speakers might produce stops with different VOT values. First, the L2 learner must tuning" may continue beyond the time that stable patterns of L1 speech production have emerged.

Production and perception could be regarded as "mismatched" if, for example, an individual produced /t/ with a mean VOT value of 25 ms, but labelled stimuli with a 25-ms VOT value as /d/. Evidence from a study which examined VOT in both speech production and perception suggests that such mismatches become less frequent for monolingual English speakers with increasing age. Zlatin and Koenigsknecht (1975, 1976) found that English-learning 2-year-olds produced /p,t,k/ with VOT values matched to their voiceless stop category 64% of the time, compared to 90% for 6-year-olds and 97% for adults. Similar results were reported for children 20 to 30 months of age by Bailey and Haggard (1980).

One possible inference to be drawn from this finding is that evolving norms of perception shape stop production (Lisker & Abramson, 1970). However, data presented by Bailey and Haggard lend themselves to an alternate interpretation, namely that individuals develop control of voiced and voiceless stops independently rather than learning a single phonological opposition as supposed by Jakobson (1968). If so, production may be influenced by factors external to evolving perceptual norms. For example, some children's early tendency to produce primarily short-lag stops may follow from neurophysiological constraints (Kewley-Port & Preston, 1974) rather than from their attempt to reproduce what they have previously heard and stored in phonetic category prototypes.

The existing VOT data, however, is generally consistent with the view that changes in perception precede corresponding changes in production. For example, the tendency of some children to briefly "overshoot" the phonetic norms of English for VOT in the production of /p,t,k/ (see Macken, 1980; Suomi, 1976) is unlikely to precipitate a change in the perception of those stops. Bailey and Haggard's (1980) data showed that children in the age range where the overshoot phenomenon is most likely to occur all showed similar VOT values, rather than showing the marked differences one would expect if individual children were experiencing a sudden marked change in perception. This leads to the inference that overshoot in speech production corresponds to a child's developing awareness of the phonetic importance of the VOT dimension (which might admit a sudden, cognitive reorganization), or from the relatively sudden discovery of an articulatory means for achieving a desired phonetic end.
rely on an auditorily-derived representation of the L2 phonetic norm (or prototype) to guide stop production in L2 since differences in laryngeal timing are evident auditorily but not visually. If the L2 learner fails to perceive VOT differences between similar L1 and L2 stops, he or she might not be able to produce L2 stops authentically because of insufficient knowledge concerning how L2 stops “ought” to be produced. Second, the L2 learner might not be able to produce L2 stops authentically because of difficulty in modifying previously established (LI) patterns of segmental articulation, even if he or she does perceive the VOT difference between L1 and L2 stops accurately. Before we consider studies examining naturalistic L2 acquisition and laboratory training studies, we will first consider a priori evidence concerning the modifiability of the VOT dimension in stop production and perception.

If L2 speech were phonologically filtered or perceived categorically, the L2 learner would not be expected to perceive VOT differences between similar L1 and L2 stops. While phonological filtering and categorical perception might contribute to the maximally efficient processing of LI sounds, it might be viewed as nonadaptive for the perception of L2 if it prevents the learner from perceiving acoustic differences between similar LI and L2 sounds.

There is reason to question the importance of phonological filtering, but listeners frequently do judge acoustically different sounds as belonging to the same phonetic category. For example, when /p,t,k/ is produced with short-lag rather than the customary long-lag VOT values by deaf speakers of English (Monsen, 1976), English-learning children (Eilers & Oller, 1976; R. Eilers, personal communication), and adult non-native speakers of English (Lisker, 1978; Elsendoorn, 1983a, 1983b), these stops are often heard as voiceless stops by native English-speaking listeners. However, the evidence reviewed earlier suggested that L2 learners probably perceive or at least detect VOT differences between similar LI and L2 sounds.

Previous research has shown that within-category discrimination can be improved by training (Carney, Widen, & Viemeister, 1977; Soli, 1983), so it is not surprising that laboratory training studies aimed at teaching subjects to identify a new category of stops have been successful. In one study (Lisker, 1970), discrimination training enabled native speakers of Russian, a language with lead and short-lag stop categories, to label synthetic short- and long-lag stimuli as /bα/ and /pa/). Strange (reviewed in Strange & Jenkins, 1978) attempted to train English speakers to identify members of a lead category in a continuum of CV syllables synthesized with VOT values ranging from -150 to +150 ms. Native English speakers typically hear only two categories in this stimulus range. Six hours of training with either immediate or delayed feedback resulted in only limited improvement in subjects’ ability to discriminate lead and short-lag stops, and training did not generalize to untrained places of articulation.

Strange (Strange & Jenkins, 1978, p. 152) concluded that “changing the perception of the VOT dimension . . . is not easily accomplished by techniques that involve several hours of practice.” However, a later study (Pisoni et al., 1982) showed that native English-speaking subjects could be trained to identify three categories of stops through the alternating presentation of “good” exemplars of the three categories (i.e., stimuli with -70, 0, and 70 ms VOT values; see also McCloud, et al., 1983). Other studies have shown that it is possible to train non-native adults (Kalikow & Swets, 1972), deaf individuals (Stark, 1972), and aphasics to perceive the English VOT distinction between stops.

Taken together, the evidence from the studies just cited suggests that individuals can learn to modify their use of VOT in labelling prevocalic stops. If several hours of laboratory training may enable subjects to identify a new category (e.g., Pisoni et al., 1982) or label stops differently (e.g., Kalikow & Swets, 1972), it seems reasonable to think that L2 learners will also be able to learn to perceive the phonetic difference between LI and L2 sounds differing in VOT during naturalistic L2 learning.

This raises the question of whether L2 learners have the capability of producing what they perceive. The Yeni-Komshian et al. (1968) study reviewed earlier indicated that adults may be unable to imitate stops accurately if they have VOT values atypical of stops in LI. Weismer (1980) suggested that the time course of the laryngeal devoicing gesture used to suppress voicing in /p,t,k/ is “pre-programmed,” presumably as the result of learning which occurs during LI acquisition. If so, then perhaps the duration of VOT (if not temporal onset of the laryngeal devoicing gesture) is not under volitional control. Talkers might conceivably lose the ability to modify laryngeal timing once patterns of speech motor control that are optimally specified for LI production have been established. However, based on their cross-language work, Lisker and Abramson (1970) concluded that there is no absolute physical limit on talkers’ ability to control laryngeal timing (and thus VOT).

The results of a laboratory training study by Weismer and Cariski (1983) supported the belief that listeners can modify VOT. Two adults were instructed to produce the stop in /ata/ with VOT values that equaled the average VOT measured in their normal production or were 30 ms longer or 30 ms shorter. The subjects were given feedback after each attempt at reproducing a target VOT value (“correct,” “high,” or
“low”), and their success in achieving the externally specified target was noted. The subjects were generally successful in producing /t/ with longer or shorter VOT values, as required. Their average (unsigned) errors were about 10–15 ms, with no clear pattern of undershooting the longer-than-average VOT target or undershooting the shorter-than-average VOT target.

Another study using a different paradigm (Flege & Hammond, 1982) also supported the belief that LI patterns of laryngeal timing are modifiable. Native English-speaking subjects were required to read English sentences with what they considered to be a “typical” Spanish accent. The authors hypothesized that English speakers familiar with Spanish-accented English would be able to produce stops with the VOT values typical of Spanish-accented English (i.e., with VOT values shorter than those produced by native speakers of English). The subjects produced /t/ with VOT values that were significantly shorter than those observed in the speech of native English subjects producing the same speech material without special instructions. The subjects did not merely produce /t/ with VOT values appropriate for an English /d/ (i.e., /d/ for /t/). Instead, they produced /t/ with VOT values which ranged from the short-lag range typical of Spanish to the long-lag values typical of English.

In summary, the studies just reviewed suggest that L2 learners possess the ability to perceive accurately the acoustic differences between similar L1 and L2 stops and to modify VOT in speech production. L2 learners should therefore be able to modify VOT in such a way as to produce L2 stops authentically. However, we have discussed factors such as the tendency to maintain LI production patterns and the slow process of learning which hinder the L2 learner’s progress in learning to produce L2 stops authentically. We next turn to studies of naturalistic L2 learning in order to learn how successful L2 learners have been in producing L2 stops having VOT values which differ from VOT values seen in L1.

VOT in L2 Speech Acquisition.

Recent instrumental studies have amply confirmed Daniel Jones’ observation (1948) that non-native speakers of English tend to realize English stops with patterns of laryngeal timing typical of their L1. Two studies showed that native speakers of Arabic produced stops in English words with VOT values comparable to those measured in Arabic words. As mentioned earlier, Arabic /t/ and /k/ are “long-lag” stops produced with somewhat shorter absolute VOT values than their English counterparts (Flege, 1979). Thus the task facing native Arabic-speaking learners of English is to modify Arabic patterns of laryngeal timing so as to produce a somewhat longer VOT interval. Flege and Port (1981; see also Flege, 1980) examined the production of English /p,t,k/ by two groups of Saudi Arabians differing in English-language experience. These L2 learners did not modify VOT in producing English stops. Their mean VOT was significantly shorter in English words (about 31 ms) than that of monolingual English speakers (56 ms) and did not differ significantly from the VOT measured in their production of comparable Arabic words. Port and Mitleb (1983) found that two groups of Jordanians differing in English-language experience produced English /p/ and /t/ with VOT values that were significantly shorter (by 35 ms) than that of native English speakers.

One possible explanation for why the adult native Arabic learners of English did not seem to modify VOT in producing English stops is that the VOT difference between similar stops in Arabic and English is relatively small. Other studies have examined the speech of individuals faced with the need for increasing VOT to an even greater extent in L2 than native speakers of Arabic. Native speakers of Spanish and French must learn to produce /p,t,k/ as long-lag rather than short-lag stops. These studies have indicated that L2 learners may produce stops with different VOT values in L1 and L2 words when the VOT values in L1 and L2 stops differ substantially.

Two studies examined the English spoken by native speakers of French. Caramazza et al. (1973) measured the prevocalic stops spoken by French Canadian students who began learning English at about the age of 6 years. They produced /p,t,k/ with significantly longer VOT values in English words (51 ms) than French words (28 ms) but with shorter values than those observed in the production of English words by monolingual native speakers of English (74 ms). Flege and Hillenbrand (1984) found that French women who had learned English as adults and lived in the United States for 12 years produced English words with shorter mean VOT values (55 ms) than native English speakers (75 ms) in three different speaking tasks. However, unlike the French subjects in the Caramazza et al. (1973) study, they did not produce stops with longer VOT values in English than French words.

Williams (1977b) measured stops produced by adults who began learning their L2 (Spanish or English) by about the age of six years, and whose accent in both L1 and L2 was deemed native-like. Williams reported that her bilingual subjects produced /p,t,k/ in Spanish words with VOT values that did not differ from those of monolingual native speakers of Spanish. They also produced /p,t,k/ in English words with
VOT values that did not differ significantly from those of monolingual native speakers of English. Mack (1984) measured /t/ in English words produced by native French adults who lived in the United States, had learned English before the age of eight years, and were judged by themselves and other native speakers of French to pronounce English better than French. Her bilingual subjects produced the /t/ in English words with mean VOT values that did not differ significantly from those of age-matched English monolinguals. (The subjects’ production of /t/ in French words was apparently not measured.)

The findings of Williams (1977b) and Mack (1984) appear to differ from results reviewed earlier. They suggest that certain L2 learners, perhaps those who are highly gifted for L2 pronunciation, have begun to learn L2 at an early age, or speak L2 to the near exclusion of L1, may successfully modify VOT in producing stops in L2. However, there is reason for caution in accepting any of these conclusions. Mack (1984) did not provide information pertinent to her subjects’ previous English L2 or French L1 experience or current use of L1 and L2. The subjects examined by Williams (1977b) included individuals whose L1 was either Spanish or English. The lack of a statistically significant difference between the bilingual subjects in Williams’ study and English monolinguals may conceivably have been due to the production of the subjects whose first language was English. (This cannot be ascertained, since the report did not include measures of individual or group means, nor measures of central tendency for the monolingual and bilingual groups.) VOT was measured in imitated rather than spontaneously produced speech. Although children appear to produce the same VOT values in both conditions (Bond & Korte, 1983), it might potentially affect the speech of adults who are aware of VOT differences between sounds found in L1 and L2 (see Elman et al. 1977).

This caution appears warranted in light of Williams’ later findings (1979, 1980) for children whose native language was Spanish. Puerto Rican children produced /p/ with longer VOT values in English than Spanish words, indicating awareness of the VOT differences distinguishing voiceless stops in those languages. However, even those children who arrived earliest (8 to 10 years of age) and had lived longest in the United States (3 to 3.5 years) seem to have produced English /p/ with shorter VOT values than age-matched English-speaking children. The evidence reviewed earlier makes it seem unlikely that adults will outperform children in producing a foreign language, at least beyond the initial stages of L2 learning.

Kewley-Port and Preston (1974) hypothesized that it is more difficult to produce long-lag than short-lag stops. It might therefore be easier to learn to produce short-lag than long-lag versions of /p,t,k/ in an L2.

This hypothesis was supported by Flege (1985; see also Flege & Fillenbrand, 1984). Native English speakers who had spent a year in Paris produced the /t/ in French words with significantly longer VOT values (72 ms) than French monolinguals (33 ms). However English native speakers who held advanced degrees in French and had lived in France for about one year, and English native speakers who had lived for 12 years in Paris, produced /t/ in French words with VOT values that were non-significantly longer (about 45 ms) than French monolinguals.

Taken together, these studies suggest that L2 learners often approximate the phonetic norms of L2 by modifying L1 VOT values, but seldom match L2 native speakers. All of the studies reviewed thus far have examined the production of similar rather than new stops in L2. One might hypothesize that it is easier for L2 learners to produce L2 stops with accurate VOT values if the L2 stops being learned have no obvious counterparts in L1. However, Flege and Port (1981) found that Arabs produced /p/ with significantly shorter VOT values than native English speakers. Since there is no /p/ in Arabic, it might be regarded as a new stop for Arabs (although they may possibly regard it as a “voiceless /b/”).

Such an argument cannot be used to explain the results of Suomi (1980) concerning the production of English stops by Finns. English /p,t,k/ must surely be regarded as new sounds for Finns because in most contexts Finnish has only a single series of short-lag stops (written “p,” “t,” “k”). Suomi (1980) found that Finnish high school students and teachers produced /p,t,k/ in English words with shorter VOT values than native English speakers. They produced /p,t,k/ in English words with Finnish-like short-lag VOT values more frequently than native speakers of English (18% versus 5%).

Data reported by Suomi (1976), on the other hand, seemed to support the hypothesis that L2 learners are able to produce new L2 stops with accurate VOT values. Finns majoring in English at the university produced English stops with long-lag VOT values (M = 62 ms; SD = 30) similar to those in stops produced by native English speakers (M = 59 ms; SD = 17). The Finnish subjects produced no overlap in VOT values between voiced and voiceless stops. These results must be interpreted cautiously, however. The speech material consisted of phrases read from lists. Some of the Finns seem to have overshot the long-lag English phonetic norm for /p,t,k/ in a way that is reminiscent of children acquiring English as an L1 (see Macken, 1980). It is possible that the grouped values reported for the Finnish and English native speakers did not differ because the means for Finns who “overshot”
was pooled with the VOT values of Finns who “undershot” English /p,t,k/ (i.e., produced these stops with VOT values that were shorter than values produced by the native English speakers).

In summary, L2 speech production studies have indicated that relatively inexperienced L2 learners often produce L2 stops with inappropriate VOT values that closely resemble values for similar stops in L1. Inexperienced L2 learners seem to produce L2 stops with VOT values that are too short for L2 when /p,t,k/ are realized with shorter VOT values in L1 than L2; and with VOT values that are too long for L2 when the VOT of /p,t,k/ is longer in L1 than L2. L2 learners who are relatively more experienced have been observed to modify the VOT value in similar L2 stops. However, they usually seem to produce L2 stops with “compromise” VOT values (Williams, 1980) which are intermediate in value to stops produced by monolingual native speakers of L1 and L2. It appears that even highly experienced L2 learners seldom match L2 native speakers in producing the VOT in similar L2 stops if they have begun learning L2 after about the age of six years. The possibility exists that experienced L2 learners who have begun learning L2 at a later age may match L2 native speakers in producing L2 stops which have no direct and obvious counterpart in L1 (i.e., new sounds).

### VOT in L2 Perception.

It seems reasonable to ask whether the apparent failure of L2 learners to produce L2 stops with accurate VOT values follows from speech perception. L2 research indicates that although perception of L2 stops may evolve during the process of L2 learning, L2 learners often differ from native speakers of L2 in perceiving stop continua differing in VOT.

Caramazza et al. (1973) provided evidence that the perception of stops changed as the result of L2 learning. French speakers of English labeled synthetic CV stimuli differing in VOT in a different way than monolingual speakers of French or English. Their phoneme boundaries for /b-p/, /t-d/, and /k-g/ continua occurred at values that were intermediate (averaging 23 ms) to those observed for monolingual French (14 ms) and English (29 ms) subjects.

Caramazza et al. (1973) explored the issue of whether French speakers of English had different perceptual standards for stops in French and English by presenting the same continua of CV stimuli in a “French” and “English” mode. They found that the French Canadian speakers of English labeled stimuli the same way in both language sets. Williams (1977b) obtained similar results in examining the perception of individuals who spoke Spanish and English proficiently. The experimenter spoke either Spanish or English before and during the experiment in order to vary language set.

Elman et al. (1977) later provided evidence that the listener’s language set affects perception. This study made use of naturally produced (rather than synthetic) stimuli inserted into a Spanish or English carrier phrase. The stimuli were interspersed with naturally produced Spanish or English words, depending on the language set. The stimuli of interest were stops with short-lag values of 15 and 26 ms. English monolinguals heard the stimuli as /b/, whereas monolingual Spanish speakers clearly heard them as /p/. Native Spanish speakers with good English pronunciation, but not those whose English pronunciation was less authentic, labeled short-lag stops as /b/ significantly more often in the English than Spanish set. The finding of Elman et al. (1977) thus suggested that highly proficient L2 learners may develop separate phonetic category prototypes for similar L1 and L2 sounds.

This conclusion is supported further by evidence that individuals who speak two languages may maintain sensitivity to the acoustic difference distinguishing stops in L1 while developing sensitivity to the acoustic difference distinguishing stops in L2. Data reported by Caramazza et al. (1973) indicate that the percentage of voiceless stop responses given by native French subjects did not increase monotonically along with increases in VOT as it did for native English subjects. There were increases in the frequency of voiceless responses at around 0 ms, that is, near the phoneme boundaries expected for French monolinguals. However, this response increase was not interpreted as evidence for a phoneme boundary because it did not represent a crossover from predominantly voiced to voiceless stop judgments as it did later in the continuum (i.e., near the English phoneme boundary). The bimodal pattern was evident for all native French subjects, but more so for French monolinguals than French speakers of English.

Williams (1979, 1980) examined the perception of a /pa/-to-/ba/ continuum by monolingual English-speaking children and native Spanish-speaking children. Puerto Rican children aged 8 to 10 years and 14 to 16 years had phoneme boundaries at significantly shorter VOT values (8 ms and 5 ms, respectively) than the English children (20 ms). Children who had lived in the United States for three years more nearly resembled the monolingual English children (whose mean boundary was at 10 ms) than those who had lived less than 6 months in the United States (3 ms). Children in both age groups showed a peak in discriminability near the English boundary (20 ms), the magnitude of which increased with length of residence in the United States. Both groups of children showed evidence of retaining sensitivity to the acoustic contrast distinguishing voiced and voiceless Spanish stops.
older children showed peaks of discriminability in the region of the Spanish phoneme boundary established for Spanish monolingual adults (−4 ms). The younger children tended to show bimodal identification functions similar to those observed for French Canadian adults by Caramazza et al. (1973).

Two studies examined the perception of adults who learned English as children. Mack (1984) found that 10 French learners of English who pronounced English better than French did not differ from monolingual English subjects in labeling a /ta/-to-/da/ continuum. Williams (1977b) examined the perception of adults who pronounced Spanish and English with native-like proficiency. Three showed a phoneme boundary near the Spanish monolingual boundary of −4 ms, the remaining five a phoneme boundary near that of monolingual native English speakers (25 ms). In a discrimination test, the subjects with a Spanish-like phoneme boundary showed enhanced sensitivity to pairs of stimuli straddling both the Spanish and English phoneme boundaries. In this respect, they resembled the older Puerto Rican children mentioned earlier. Subjects with an English-like phoneme boundary, on the other hand, showed enhanced discrimination only for pairs of stimuli which straddled the English phoneme boundary. This last finding suggests the possibility that as L2 learners acquire sensitivity to new acoustic distinction, they may lose sensitivity to the acoustic dimensions distinguishing stops in L1.

In summary, the findings just reviewed suggest that perception evolves slowly as L2 learners encounter the phonetically different sounds of L2, but not sufficiently to allow them to match L2 native speakers. The only possible exceptions appear to be the adults examined by Williams (1977b) and Mack (1984). Perhaps learning L2 by about the age of six years leads to native-like proficiency in the perception of L2 stops. Perception may differ as a function of language “set,” indicating an awareness of phonetic difference between L1 and L2 sounds. The question arises, as it did for speech production, whether a thoroughly proficient L2 learner who has learned L2 after the age of six years will ever perceive stops just like monolingual speakers of L1 and L2. Even the adults in the Williams' (1977b) study perceived stimuli either like Spanish or English native speakers. Thus we must tentatively conclude that individuals who speak two languages do not function just like monolingual native speakers of L1 and L2 when perceiving their two languages (see Mack, 1984).

Perhaps the most interesting result of the VOT studies reviewed here is that the same L2 learners who approximated L2 phonetic norms for VOT in production also showed evidence of changing perception: their phoneme boundaries were intermediate to those of monolingual native speakers of L1 and L2 (Caramazza et al., 1973; Williams, 1977b, 1979, 1980; cf. Albert & Obler, 1978; Obler, 1983). This suggests that production and perception evolve together. Many researchers believe that the changes in perception which occur in speech acquisition or learning “lead” corresponding changes in production. In the next section we will consider reasons why L2 learners seem to be destined to produce and perceive L2 sounds with VOT values intermediate to those observed for monolingual native speakers of L1 and L2.

**EMERGING ISSUE AND THEORIES**

The research reviewed here indicated that L2 learners, adults as well as children, seldom match monolingual native speakers of a target L2 in producing and perceiving L2 sounds. It is nonetheless true, however, that many experienced L2 learners at least approximate the phonetic norms of L2. That is, they produce and perceive L2 sounds differently than similar L1 sounds as the result of their L2 experience. Previous research has largely focused on explaining why L2 learners differ from L2 native speakers. However, there is just as much need for explaining the partial approximation that is observed in L2 speech production. That is the focus of this section.

**PHONOLOGICAL TRANSLATION**

It is frequently claimed that foreign accent is inevitably present in adult L2 speech because adults have passed a “critical” or “sensitive” period for speech learning by the time L2 learning commences. However, as Oyama (1979) noted, the critical period hypothesis does not represent an explanation of foreign accent. Instead, it seems that many differences between native and non-native speakers derive from the maintenance of L1 patterns for the production and perception of L2 sounds. An L2 sound may be replaced by an L1 sound with which it has been identified. Such substitutions result in authentic pronunciation of L2 sounds that are identical acoustically to sounds in L1, but they result in the nonauthentic pronunciation of what we have called “similar” and “new” L2 sounds.

Experiments (e.g., Flege, 1984a) have shown that L2 learners are able to detect acoustic differences between similar L1 and L2 sounds. However, it is possible that in early stages of L2 learning all sounds encountered on the phonetic surface of L2 (whether they are identical, similar, or new) are identified with L1 sounds. The L2 learner may not attend to subcategorical acoustic differences distinguishing L1 and L2 sounds in early stages of L2 learning because of the compelling need to understand what is being said. Perhaps it is only later that the L2 learner ceases to identify similar and new L2 sounds with counterpart
sounds in LI. The need to perceive sounds efficiently and to produce L2 sounds authentically might favor a "code forming" as opposed to "code using" strategy in later stages of L2 learning (Redmond, 1977) once the learner has become familiar with the lexicon, syntax, and phonology of L2.

L2 learners may eventually establish a central phonetic representation for a new L2 sound once they have ceased to identify it with a sound in LI. If so, any divergences from L2 phonetic norms that exists in their realization of the L2 sound might simply be the result of a learning process that has not yet reached completion. Learning to pronounce new L2 sounds authentically takes time, just as mastering the production of sounds in LI acquisition.

This same explanation might not apply to similar L2 sounds that are produced non-authentically. Previous studies suggest that L2 learners may approximate only partially the phonetic norms of L2 for similar L2 sounds even after many years of L2 experience. For example, Flege (1985) found that French learners of English, and English learners of French, continued to produce /t/ with VOT values differing from those observed for monolingual native speakers of L2 after having lived for some 12 years in an L2-speaking environment. Children acquiring L1 ordinarily establish native-like control of VOT in L1 stops by the age of five or six years. It is therefore unlikely that such divergences from the phonetic norms of L2 by adult L2 learners is evidence of "learning in progress."

One might argue that adults' partial approximation of L2 phonetic norms for similar L2 sounds is a consequence of decreases in the frequency of interlingual identifications. An L2 learner might appear to approximate partially L2 phonetic norms in producing a similar L2 sound if the frequency with which it is identified with an L1 counterpart sound was diminishing. Such a learner might substitute an L1 sound for the similar L2 sound part of the time, and directly attempt the L2 sound without reference to previous habits the remainder of the time. The problem with this explanation is that it is based on the assumption that perception of the degree of similarity between pairs of sounds in L1 and L2 may change rapidly. The implication is that at one moment the L2 learner perceives sufficient phonetic similarity between a pair of sounds in L1 and L2 to substitute one for the other, but soon thereafter perceives sufficient dissimilarity not to do so.

A more plausible possibility is that the degree of perceived similarity between similar L1 and L2 sounds decreases over time. Unfortunately, I know of no existing study which has directly tested this hypothesis. It might be tested, for example, by examining the perception of Dutch learners of English. According to analyses by Disner (1983), the /i/ of Dutch and English are acoustically identical sounds, whereas the /i/ of Dutch and English differ acoustically and are probably what we have been calling "similar" sounds. If the degree of perceived similarity between pairs of similar L1 and L2 sounds decreases as the L2 learner becomes familiar with the acoustic differences between them, then Dutch learners should perceive greater dissimilarity between the Dutch and English /i/ as they gain experience in English. There should be no change over time in perceived similarity, on the other hand, for the identical Dutch and English /i/.

L2 learners may develop distinct perceptual "targets" for similar sounds in L1 and L2 as they become aware of the acoustic differences distinguishing pairs of similar sounds in L1 and L2. Three studies have tested the hypothesis that speech sounds in L1 and L2 are perceived differently. One (Elman et al., 1977) provided evidence that at least highly proficient bilinguals label sounds differently in their two languages. However, assuming that L2 learners develop distinct perceptual targets for similar L2 and L1 sounds, we are still left with the need to explain why they do not produce the L2 sound authentically.

Flege (1981; Flege & Hillenbrand, 1984) postulated that L2 learners actively integrate their phonetic experience hearing and speaking L2 sounds with their prior experience hearing and speaking L1 sounds judged to be equivalent. He proposed that L2 learners merge the phonetic features of similar L1 and L2 sounds as a result of equivalence classification and that, as a result, they succeed only partially in approximating L2 phonetic norms for similar sounds in L2.

This hypothesis rests on several important assumptions. The first is that adults do not differ from children in terms of a basic sensory-motor ability to translate the patterns of auditory and visual stimulation associated with segmental articulation into stable articulatory motor plans. The second assumption is that central representations for speech sounds continue to be influenced by phonetic experience beyond the establishment of phonetic categories in L1. The third assumption is that the accuracy of central representations limits the authenticity with which L2 sounds may be produced.

Equivalence classification is clearly needed by children for acquiring L1. It would probably be impossible for a child to learn speech if, for example, the acoustically different tokens of /i/ produced by males and females were not identifiable as belonging to the same category. During L1 acquisition, the child establishes perceptual constancy for sound categories by identifying a wide range of acoustically different sounds as categorically equivalent (Kuhl & Miller, 1982). The young child is able to attend to just those sensory attributes or features of a sound important to its category identity while learning to ignore other attri-
butes that are irrelevant, less important, or which provide conflicting information concerning categorical identity (Kuhl, 1980; Tversky & Gati, 1978).

There is reason to suppose that adults resemble children in their ability to sort out relevant and irrelevant features of speech sounds (e.g., Hillenbrand, 1983). However, since L2 learners are typically older than children acquiring L1, it is likely that most L2 learners differ from children acquiring L1 in terms of the number of categories they have established, and according to how thoroughly central representations have been elaborated. A child just beginning to acquire L1 probably has relatively few, only partially elaborated, prototypes. For such a child, many if not all L2 sounds must be regarded as new sounds. The adult L2 learner, on the other hand, approaches L2 with a full complement of fully elaborated prototypes and may therefore regard few if any L2 sounds as being new sounds.

Flege (1981) proposed that L2 learners seek and find perceptual constancy in the sounds they encounter in L2 and L1 because they judge acoustically similar sounds in L1 and L2 to be categorically equivalent, even though they may detect the acoustic differences distinguishing them (Flege, 1984a; Flege & Hammond, 1982). Because of this, the central representations developed by L2 learners for sounds in L2 will not match those of monolingual native speakers of L2, whose perceptual representations are based on their experience with L2 sounds only.

We noted earlier the close relationship that exists between specific measures of speech production and perception for mature monolinguals (e.g., the average VOT value measured for /p,t,k/, and the point along a VOT continuum where judgments shift from voiced to predominantly voiceless). Previous L1 research (e.g. Strange & Broen, 1980) suggests that there is often no clear or simple relation between specific measures of perception and the development of control over various aspects of segmental articulation. However, the extent to which measures of children’s speech perception resemble that of adults often exceeds the extent to which specific aspects of their speech production resemble adults’.

Existing evidence suggests that there is also a close match between specific aspects of L2 speech production and perception by experienced adult learners of an L2 (Caramazza et al., 1973; Williams, 1980). Flege (1981; Flege & Hillenbrand, 1984) proposed that the two are related. Specifically, the accuracy with which a similar L2 sound is produced is limited by the extent to which perception of that sound has evolved as the result of L2 experience. This hypothesis presupposes that the information in central representations responsible for speech sound identification and articulation are functionally separate and nonisomorphic. This assumption receives support from the finding that L2 learners are often conscious of their own pronunciation mistakes (e.g., Neufeld, 1980).

The limiting effect of perception on production might come about because L2 learners develop central representations for L2 sounds differing from those of monolingual speakers of L2. It seems reasonable to think that the central representations developed by monolinguals depend importantly on the many tokens of an L1 sound they have processed phonetically. It is hypothesized that the central representation developed by L2 learners for a similar sound in L2, on the other hand, is based on hearing many tokens of the sound in L2 and on many the tokens of an L1 sound(s) judged to be equivalent to it.

The two-language source of input for L2 sounds can be illustrated by considering the /t/ of French and English. French monolinguals generally hear /t/ produced with short-lag VOT values whereas English monolinguals hear /t/ produced with mostly long-lag VOT values. French and English monolinguals will experience a unimodal distribution of VOT values for /t/, whereas the French learner of English, and the English learner of French, will experience a bimodal distribution of /t/ tokens. As a result, any measure of central tendency for bilinguals will gradually begin to shift away from monolingual values in the direction of L2 values. However, given the vast number of L1 stops experienced prior to L2 learning, the central tendency is likely to remain distinct from that of L2 native speakers.

Oden and Massaro (1978) proposed that speech perception is mediated by category prototypes containing probabilistic information concerning the central tendency of various acoustic dimensions. Flege (1981) proposed that prototypes also mediate the perception of L2 sounds. Assuming that prototypes are based on experience and change as the result of experience beyond L1 acquisition, the prototype for English /t/ developed by French learners, and the prototype for French /t/ developed by English learners, will specify a VOT value intermediate to that of French and English monolinguals. If so, L2 learners will not perceive /t/ in their L2 just like monolingual native speakers of L2. Assuming further that production is eventually modified until it conforms to information specified perceptually in the prototype, L2 learners are destined to merely approximate the phonetic norms of L2 for similar L2 sounds, at best.

The hypothesis that limits on changes in perception in turn limit changes in production was supported by a factor analytic study (Snow & Hoefnagel-Höhle, 1979). This study examined several aspects of Dutch L2 learning by native English speakers: spontaneous and imitative
pronunciation; fluency; auditory discrimination, aural comprehension, vocabulary, syntax, and grammaticality judgments. Pronunciation did not stand out as a clear factor after several months of naturalistic L2 learning. After seven months there was a weak correlation between measures of speech production and perception, which loaded onto separate factors. By the end of one year, a strong correlation between measures of production and perception had emerged. The analysis yielded a pronunciation factor distinct from a factor related to grammatical aspects of language competence. The authors concluded (1979, p. 159) that “perceptual ability (eventually becomes) the limiting factor in pronunciation skill.”

The hypothesis that limits on changes in perception limit the accuracy of L2 speech production (Flege & Hillenbrand, 1984) leads to the expectation that whereas L2 learners’ perception may sometimes resemble that of L2 native speakers to a greater extent than related measures of their speech production, the reverse should never occur. Existing evidence generally supports this prediction, although some L2 data suggest the opposite (e.g., Goto, 1971). It is of course impossible to prove the null hypothesis that L2 learners will never perceive L2 sounds exactly like L2 native speakers because they have a two-language source of phonetic input for similar L2 sounds. However, existing evidence is consistent with the prediction that the perception of L2 learners differs from that of L2 native speakers. For example, experiments examining perception of the voicing feature in stops differing in VOT show that the phoneme boundaries of L2 learners fall at values intermediate to those observed for L1 and L2 monolinguals (Caramazza et al., 1973; Williams, 1980; but cf. Obler, 1983).

The most important prediction of Flege’s (1981) model is that L2 learners will never match monolingual native speakers of L2 because the prototypes they develop for L2 sounds differ from those of L2 native speakers. This prediction cannot be proven directly but it is supported by the existing evidence. For example, Mack (1984) found that native French speakers who were highly proficient in English pronunciation labeled fewer members of an /i/-to-/I/ continuum as /i/ than English monolinguals. These bilingual subjects did not perceive /i/ like just English monolinguals even though they had learned English early and well. The production data reviewed earlier indicated that whereas L2 learners generally do not match L2 phonetic norms for VOT and vowel timing, they may do so insofar as the duration of stop closure intervals is concerned.

Another important prediction of Flege’s (1981) model is that learning L2 will affect the production and perception of similar sounds in L1. If the phonetic properties of L1 sounds influence the prototypes developed for similar L2 sounds as the result of equivalence classification, the reverse should also occur. This prediction was supported by a study (Elman et al., 1977) examining the perception of stop consonants by Spanish learners of English. Monolingual English speakers heard naturally produced stops with VOT values of about 20 ms as /b/, whereas monolingual Spanish speakers heard them as /p/. Spanish/English bilinguals heard considerably more of the short-lag stops as /b/ than the Spanish monolinguals.

Several existing studies supported the prediction that learning L2 affects the production of sounds in L1. The results of Caramazza et al. (1973) indicated that the VOT in French words spoken by French speakers of English were slightly longer (hence English-like) than French stops spoken by French monolinguals. Williams’ (1979, 1980) study revealed that Puerto Rican children who had learned English produced Spanish /p,t,k/ with somewhat longer VOT values (about 30 ms) than would be expected for monolingual Spanish-speaking youngsters (0–10 ms). Finally, Flege (1985) found that native French women who had lived in Chicago for 12 years, and native English women who had lived in Paris for 12 years, produced the /t/ in L1 words with VOT values of about 50 ms. The French women differed significantly from French monolinguals in producing /t/, and the native English women differed significantly from English monolinguals. In both instances, L1 stop production was modified in the direction of L2 phonetic norms.

In summary, equivalence classification appears to have an important influence on L2 production and perception. However, Flege’s (1981) hypothesis concerning its effect on L2 production and perception leaves at least two important questions unanswered. First, how is sensory experience with similar L1 and L2 weighted during the development of central representations for L2 sounds? And, second, do the central representations previously established for L1 sounds directly affect the central representations being developed for similar L2 sounds? Schouten (1975) speculated that there is a kind of “gravitational attraction” between L1 and L2 categories. That is, the central representation developed for an L2 sound may influence the corresponding L1 category representation. It will be important in future research to establish if equivalence classification applies only to certain phonetic dimensions such as VOT. The most important test of the hypothesis that equivalence classification limits the authenticity of L2 production will come from studies comparing the production and perception of new versus similar L2 sounds. It is predicted that experienced L2 learners may produce new sounds authentically because new sounds will ultimately evade equivalence classification, thereby permitting the learner to avoid the limiting effect of previous phonetic experience.
**PERIPHERAL SENSORY FEEDBACK**

Previous research indicates that sensory feedback is important to the establishment of central phonetic representations during L1 acquisition. However, it is less clear whether sensory feedback is used for the ongoing regulation of speech (MacNeilage, 1970) and, if so, whether it is used to the same extent by skilled talkers (e.g., mature native speakers) and unskilled talkers (e.g., children or non-native speakers).

In this section we will discuss two broad hypotheses concerning the effect of peripheral feedback on speech learning. The first hypothesis concerns the relative importance of feedback through various sensory modalities. A number of investigators (e.g., Fry, 1966) have proposed that feedback used for the regulation of speech shifts from the auditory modality to the tactile-kinesthetic modality as speech develops. The second hypothesis is that speech motor control gradually becomes less reliant on peripheral feedback as central phonetic representations are established during speech acquisition.

A number of researchers (e.g., Van Riper & Irwin, 1958; Peterson & Shoup, 1966; Siegel, Fehst, Garber, & Pick, 1980) have proposed that motor control shifts from a predominantly “closed loop” mode, in which control relies on peripheral sensory feedback, to an “open loop” mode as talkers gain experience. However, Kelso and Stelmach (1976) concluded that this view lacks solid empirical support at present, noting that “open loop” control may remain prominent in the control of the fast ballistic movements often seen in speech.

There is some support for the hypothesis that individuals rely increasingly less on auditory feedback as speech develops. Garber, Speidel, and Siegel (1980) found that masking noise had little noticeable effect on L1 speech produced by subjects as young as five years of age. However, other studies have suggested that auditory masking may have a greater affect on sounds that are controlled imperfectly than on sounds which have been mastered more thoroughly. One study (Manning, Keappock, & Stick, 1976) examined the speech of children undergoing speech therapy for an “errored” sound. The children who were relatively good in producing their errored sound in noise were more likely to maintain the gains made during therapy than those who were less able to produce their errored sound in noise.

One might interpret this finding, which has been replicated several times, to mean that the children who were able to produce their errored sound in noise controlled it better because they relied to a greater extent on “open loop” control than children who could not. Supporting this interpretation is a finding reported by Manning and Hein (1983). Adult native English speakers showed a reduced effect of masking noise on their production of L2 sounds (French /y/ and /r/) as their pronunciation of the L2 sounds improved across ten training sessions.

There also exists counter evidence to the hypothesis that reliance on auditory feedback decreases with increased skill. Siegel, Pick, & Garber (1976) reported a greater effect of sidetone amplification (i.e., a lowering of vocal output intensity) for adults than young children, but no difference in the effect of masking noise across age groups. If talkers’ use of auditory feedback decreases, or if control shifts from a primarily closed- to open-loop mode of control, the masking noise should have had a more deleterious effect on the younger than older subjects. It is of course possible that the 5-year-olds had already shifted to an open-loop mode control.

One difficulty in interpreting the results of auditory masking studies is the finding that speech production is affected no more severely by continuous speech maskers than by white noise, and that the effect of masking noise does not diminish as vocal output level is increased (Manning & Ortman, 1980). Salant-Gordon and Wightman (1983) found that continuous noise and speech maskers had equal deleterious effects on the performance of hearing-impaired and normal hearing subjects. It is possible that speech is disrupted by a masking noise simply because it distracts talkers, not because it prevents them from using auditory feedback pertinent to articulation to regulate ongoing production.

It would be useful to focus on the disruption of phonetic information in future studies of L2 speech production. Salant-Gordon and Wightman (1983) found that brief CV maskers had a greater effect on hearing-impaired than normal-hearing subjects. Similar studies should be conducted with L2 learners. For example, they could be required to produce CV syllables at regular intervals (e.g., every 900 ms) in synchrony with auditorily presented CV syllables. To test the effect of the disruption of phonetic information, syllable production could be masked by CV syllables that either matched acoustically or did not match the CV syllable being produced. If L2 learners make use of auditory feedback relating to speech gestures, the authenticity with which syllables matching the CV masker should be less than syllables which do not match. If mature native speakers rely less on peripheral auditory feedback than L2 learners, they should show less difference between the matched and nonmatched CV maskers than L2 learners.

The effect of delayed auditory feedback (DAF) on L2 production has been examined in two studies. DAF delays the (usually amplified) airborne auditory feedback which accompanies speech articulation. Most existing evidence indicates that DAF disrupts the speech of younger children more than that of older children (see Siegel et al., 1984),
perhaps as an indirect consequence of increased speaking skill in the older children. If so, DAF might then be expected to disrupt the speech of native speakers to a lesser extent than that of non-native speakers, and advanced L2 learners less than beginning L2 learners.

Two studies which addressed this question provided divergent results. Rouse and Tucker (1966) found that DAF caused more errors (omissions, repetitions, or prolongations) in speech produced with DAF than without DAF by native-speaking subjects as well as subjects in two non-native groups. Contrary to hypothesis, DAF affected deliriously the speech of native speakers more than that of the non-native speakers. MacKay (1970) reported evidence which supported the hypothesis. DAF resulted in more disfluencies in a foreign language that the subjects did or did not speak than in the subjects' L1 (German or English). It is unclear why these two studies yielded different results, but it should be noted that both examined speech that was read rather than spoken spontaneously.

It would be useful for future L2 studies to examine the articulation of specific sounds in greater detail. A repeated measures design could be used to examine the effect of DAF on the production of both new and similar L2 sounds. Perhaps DAF will disrupt new L2 sounds more than similar L2 sounds, and perhaps the relatively great disrupting effect of DAF on new L2 sounds will disappear rapidly during early stages of L2 learning.

Orosensory acuity has been linked frequently to articulatory development (Locke, 1968; McNutt, 1977, 1979). However, like the disruption of auditory feedback, the disruption of tactile-kinesthetic feedback has generally been found to affect speech production only minimally. One study (Siegel, Gunderson, Speaks, Rocker, & Niccum, 1977) suggested that L2 sounds are affected more severely by nerve block injections than L1 sounds. This finding is consistent with the hypothesis that developing speech relies more heavily on tactile-kinesthetic feedback than established speech.

However, Borden, Harris, Fitch, & Yoshioka (1981) reported evidence providing mixed support, at best, for the hypothesis. Their study examined vowels and fricatives occurring in the subjects' L1 and in a foreign language (not necessarily spoken by the subjects). Three subjects showed a somewhat greater deterioration of L2 than L1 sounds when sensory feedback was disrupted through nerve block injections, by insertion of an oral prosthesis serving to reduce tactile feedback, and by masking noise. Four other subjects, however, showed greater deterioration of L1 than L2 sounds.

One possible explanation for the seemingly divergent results of studies examining the effect of orosensory feedback is that previous studies have examined both new and similar L2 sounds. Similar L2 sounds may be substituted by their L1 counterpart, or be realized as modified versions of the L1 counterpart. Learning a new L2 sound, on the other hand, may more nearly resemble the acquisition of L1 sounds. If so, subjects may be more apt to rely on peripheral feedback in producing new than similar L2 sounds.

In conclusion, it is unclear at present whether L2 learners rely increasingly less on peripheral sensory feedback as they gain experience in L2, whether they rely less on feedback than mature native speakers of L2, or whether the peripheral feedback that is used to produce L2 shifts from an auditory to a tactile-kinesthetic modality as L2 learning proceeds. Much further research will be needed to refine and test the two hypotheses just discussed.

ATTENTION TO SPEECH

An idea that has enjoyed relatively uncritical acceptance over the past four decades is that conscious control of speech sound articulation diminishes as L1 acquisition proceeds. Speech sounds are said to become "automatized" gradually. This view is consistent with the fact that activities requiring coordinated movement generally seem to require less and less concentration and effort as skill is established (Klein, 1976). Abbs and Cole (1982, p. 174 ff.) concluded that "consciously encoded, afferent triggered, motor subroutines" may play a crucial role in the coordination of the multi-articulator movements of speech, but that such control processes are eventually "relegated to the level of unconscious or 'automatic' mediation." In a similar vein, Shelton and McReynolds (1979) suggested that the control of speech shifts from higher, more "conscious" levels to lower, more "automatic," levels as speech sounds are mastered in L1 acquisition. Menn (1979) observed that children's speech seems to become less "effortful," because phonological rules become more automatic as L1 acquisition proceeds.

A corollary to the hypothesis that speech articulation becomes automatized, hence less reliant on conscious control processes, is that speech becomes less easily disruptible as speech motor control processes become established. If so, talkers differing in overall level of speech motor skills might show no apparent difference in normal speaking conditions yet manifest measurable differences when the speech production system is taxed. For example, L2 learners might differ from native speakers of L2, and advanced L2 learners may differ from beginning L2 learners, to a greater extent when distracted than when allowed to concentrate fully on their speech production.
This broad hypothesis received support from a study by Locke and Goldstein (1973), which suggested that a talker's level of vigilance is related to speech production. Children who manifested a high degree of vigilance to acoustic stimuli (an above-threshold tone interspersed between trials of a production test) not only pronounced L1 sounds better, but were also somewhat better at imitating unfamiliar L2 sounds than children who evidenced a lower degree of acoustic vigilance.

L2 learners often report that their pronunciation of L2 deteriorates when they become excited or angry (Oyama, 1982a). This might be the result of decreased "attention to speech." To my knowledge, this hypothesis has not been tested objectively. However, even if it were demonstrated that authenticity of L2 speech production diminishes in the presence of strong emotion, one could not be certain that the emotion (or the presumed consequent decrease in attention to speech) contributed directly to a decrease in L2 authenticity. This is because angry or excited speech is apt to be produced at faster-than-normal speaking rates, which might lead to an increased frequency of casual speech variants.

A related hypothesis that has yet to be tested in L2 research is that L2 authenticity decreases when the L2 learner becomes tired. The effect of fatigue could be tested by comparing the speech produced by L2 learners after a normal night's rest to speech produced following a sleepless night.

Sociolinguistic research suggests that speech sound production is affected by variations in the level of attention to speech. Monolinguals often produce the L1 sound categories referred to as "phonological variables" in several different ways, ranging from casual to more formal variants. It appears that the frequency with which the formal variants are produced varies as a function of speaking task (Labov, 1972, p. 86 ff.). For example, formal variants occur less frequently when a talker recounts a life-threatening episode than when he or she reads a list of minimally paired words in the presence of a microphone. There is some evidence that casual speech variants are learned earlier in L1 speech development than formal variants (Shockey & Bond, 1980). Perhaps L2 learners will tend to produce early-acquired variants of a sound (i.e., L1 sounds) more frequently than later-acquired variants (i.e., the L2 counterpart of an L1 sound) when attention to speech diminishes, just as monolinguals produce later-acquired (i.e., formal) variants less frequently when attention to speech is diminished by external variables.

However, three previous studies have failed to provide support for the hypothesis that decreased attention to speech results in diminished L2 authenticity. Oyama (1982a) examined the speech of Italian men who arrived in the United States between the ages of 6 and 20 years, and who had lived there 5 to 18 years. Oyama hypothesized that paragraphs read by her subjects would be judged to be more authentic than "danger of death" narratives because the emotional impact of recounting a life-threatening situation would decrease attention to speech. Contrary to hypothesis, the subjects' spontaneous speech was judged to be more authentic than speech produced in the reading task.

Flege and Hillenbrand (1984) examined the production of L2 (French and English) vowels and consonants in three speaking tasks designed to manipulate the degree of "attention to speech." The subjects read phrases from a list, made a sentence from each phrase they had just read, and finished by telling a story based on phrases from the list. If attention to speech influences L2 production, the speech from the story task should have been less authentic than the speech from the phrase-reading task. However, fine-grained perceptual analyses and acoustic measurements indicated that the effect of the speaking task on L2 production was nonsignificant.

Finally, Eckman (1981) had L2 learners produce speech in a variety of different speaking tasks. These tasks consisted of repeating isolated words, deriving one word from another (e.g., "redder" from "red"), sentence completion, and spontaneous speech. Eckman did not report a difference in segmental articulation as a function of speaking task.

Krashen (1976, 1977, 1978, 1979) has argued that L2 learners gain productive control of an L2 either through unconscious acquisition or conscious learning. He posits that "rules" pertaining to the regularities and exceptions of L2 are learned, that the learned rules are accessed only through a relatively conscious "monitoring" process, and that a learner's success in producing L2 varies as a function of the amount of time available for monitoring L2 output. What has been acquired, on the other hand, is not represented by rules and therefore does not require monitoring time to be consciously accessed. Krashen and others have argued that the L2 learner can not monitor peripheral sensory feedback pertinent to speech sound articulation because speech sounds are too brief in duration. It might nevertheless be possible for L2 learners to make less rapid adjustments by monitoring longer stretches of speech. Grosjean and Soares (1986) noted that an L1 word interjected into L2 speech is phonetically "tinged" by its surrounding environment. Switching between L1 and L2 must take some time so it is possible that general speech parameters, perhaps akin to "bases of articulation," are changed during the process of code switching.

It would be useful for future L2 research to extend Krashen's "monitor" theory to L2 speech production. Specifically, the hypothesis that L2 authenticity depends on the amount of time available to access broad speech motor parameters, phonetic realization rule, or phonological rules (see Menn, 1979) should be tested. L2 learners may be
has been directed thus far at uncovering L2 learners’ strategies. This section will propose several possible strategies that might be examined in future research.

One strategy might be stated as “Maintain a phonetic contrast between different words.” Eckman (1981; see also Tarone, 1980, 1981) reported that two native speakers of a Chinese, language without word-final stops, often produced English words like “tag” with a paragogic vowel (e.g., /təɡ/). Eckman hypothesized that they inserted a vowel in order to maintain a phonological regularity of Chinese, in which a syllable is never closed with a stop consonant. However, another interpretation is that although some of the non-native talkers had not yet learned to produce stops in word-final position (see Flege & Davidian, 1985), they nonetheless wanted to preserve the contrast between words ending in voiced versus voiceless stops. Had they chosen to delete stops rather than add a vowel, a phonetic contrast (e.g., “tag” versus “tack”) would have been lost.

Another possible strategy L2 learners might adopt is “Do not attempt sounds that are difficult to produce.” Celce-Murcia (1978, p. 38) observed that phonological avoidance might be a “general and pervasive strategy” in the speech of some children. Such a conclusion is based on frequency counts of sounds in the words children produce and the adult words they are attempting. A sound thought to be avoided is one that is attempted with a much lower-than-expected frequency. A possible strategy for avoiding a difficult sound is to not attempt words which contain it, or to use a substitute for it. For example, Ferguson and Farwell (1975) found that one child attempted many adult words containing /b/ but none with /p/ until a certain age. Schwartz and Leonard (1982) provided experimental evidence that children attempted more words made up entirely of sounds found in their phonetic inventory than words containing a sound that was not. Celce-Murcia (1978) reported that a two-year-old simultaneously acquiring English and French appeared to choose corresponding lexical items from French and English in such a way as to avoid producing “difficult” sounds. For example, the child used French couteau (/kutɔ/) instead of knife (/naɪf/), apparently to avoid producing the difficult /f/ sound.

Ferguson and Farwell (1975, p. 434) suggested that even adults may “systematically—even consciously—avoid words difficult to pronounce.” This hypothesis has apparently not been tested for adult L2 learners, but it could be. One could have L2 learners memorize word lists containing pairs of synonymous L2 words which either contain or do not contain an L2 sound known to be difficult (e.g., think, which contains the difficult /θ/ sound, as opposed to ponder). Once the entire

most successful in changing overall production parameters when they have the time and/or necessary level of vigilance to monitor their vocal output. Time may also be important to permit L2 learners to preplan upcoming speech as the result of noting discrepancies between their speech and that of L2 native speakers.

No existing study has provided a satisfactory test of the hypothesis that decreased attention to speech reduces L2 authenticity. No study reviewed earlier objectively verified that changes in speaking tasks brought about the putative changes in “attention to speech.” It would therefore be useful for future research to assess L2 speech production while the subject engages in a variety of manual tasks of varying complexity. The effect of varying the manual tasks on attention could therefore be useful for future research to assess L2, speech production.

STRATEGIES

An important assumption of much L1 acquisition research (e.g., Menn, 1980) is that the child learner seeks to discover ways to produce L1 words with the limited repertoire of sounds available. This “cognitive” approach seems to imply that the learner develops tacit metalinguistic awareness of speech units smaller than the word, and that the child learner is aware of differences between his or her speech and that of adults. The methods a child uses in producing L1 are sometimes referred to as “strategies.”

The L2 learner probably also has tacit awareness of how his or her pronunciation differs from that of L2 native speakers (Neufeld, 1979, 1980) so they may also adopt strategies for minimizing differences between their L2 speech and that of native speakers. Little research
list had been learned, the subjects would be required to use as many items as possible in spontaneous speech. Subjects might use words not containing the “difficult” sounds more frequently than words which do contain the difficult sounds if avoidance is used as a strategy for L2 production. This approach assumes, of course, that factors such as word familiarity could be controlled.

Another general L2 strategy may be “Use features that are meaningful in L1 instead of L2 features which are not.” Weinreich (1953/1963) was among the first to note that an acoustic dimension may play different roles in speech perception for individuals differing in L1 background. Earlier we discussed the phenomenon of “reinterpretation.” Research indicated that in producing the distinction between English tense and lax vowels, native speakers of various quantity languages produced a much larger temporal contrast than native English speakers. They apparently did so because they interpreted the contrast between vowels like English /i/ and /I/ as a phonemic length contrast.

Still another general strategy may be “Try to talk fluently.” L2 learners could do this, for example, by inserting hesitation vowels or prolonging vowels. Either strategy might provide added time to access the next word, or to preplan articulation. Few L2 studies have examined connected speech, and no data bearing on this issue is now available. It would, however, be a simple matter to compare the number and length of hesitation pauses in L1 and L2 speech.

Another strategy used in L2 production may be “Generalize what you have learned.” A powerful argument that speech behavior is rule governed is the observation of overgeneralization (Selinker, 1972). Native English-speaking children, for example, are observed to say “goed” instead of “went” as the result of overgeneralizing the English past tense rule.

There is some evidence that L2 learners may overgeneralize what they have learned about producing L2 sounds. Walz (1979) observed that first-year French students who had previously studied Spanish used a trilled /r/ in place of the uvular /r/ of French. F. Johansson (1973) observed that adult L2 learners used a new L2 (Swedish) sound they had learned for another new L2 sound they had not yet mastered. My own observation is that Saudi Arabians substitute the similar /b/ of Arabic for the new /p/ in English words, and also sometimes realize /b/ in such a way that it is heard as /p/ (Flege & Port, 1981). Some substitutions like these may of course stem from the incorrect lexicalizations of L2 words, but they may also indicate L2 learners’ awareness that certain sounds are “difficult.” That is, the L2 learners may be attempting to generalize previous articulatory solutions to difficulties not yet overcome.

**CONCLUSION**

This chapter has reviewed many studies examining the production and perception of sounds in a foreign language. It should be obvious that the factors affecting L2 pronunciation are numerous and interconnected. They include cognitive, social, and psychological factors as well as the sensory and motoric aspects of speech learning. In most instances we do not understand these factors taken singly, much less how they interact with one another. A great deal of interdisciplinary research will be required to test hypotheses and develop models relating to the process of L2 learning.

Continued efforts should be directed at exploring the extent to which adult L2 learning differs from child L1 acquisition, as well as to compare the L2 learning of children and adults. Research should also focus on L2 learners’ ability to switch between L1 and L2 modes of phonetic implementation and perception; the relationship between changes in production and perception of L2 sounds; the perception of similarity between pairs of sounds in L1 and L2; the role of sensory input; and the role of maintenance (i.e., interference) in various phases of L2 learning.

Another major goal of future research should be to determine the role of phonological structure in speech learning. Underlying the production and perception of the speech sounds in L1, and presumably L2, are representations and categories that are more abstract than the sensorimotor codes used in phonetic implementation and phonetic perception. The nature of the interlocking network of contrasts the learner brings to L2, that is, the phonology of L1, may importantly affect more basic processes of speech learning. In attempting to answer these questions, researchers will contribute to a better understanding of a fundamental and very important human ability.

**REFERENCES**


to both perception and production research in second language acquisition. Research in Phonetics (No. 2, pp. 227-260). Department of Linguistics, Indiana University, Bloomington, IN.


Elsendoorn, B. A. G. (1980). Vowel duration as a cue to a Dutch foreign accent in English CVC-words. PRIPU (No. 5, pp. 64-86), Institute of Phonetics, University of Utrecht, The Netherlands.


Garnes, S. (1977). Effects of bilingualism on perception. In O. Garnica (Ed.), *Papers in psycholinguistics and sociolinguistics* (No. 22, pp. 1-10). Department of Linguistics, The Ohio State University, Columbus, OH.


Hanlon, C. (1971). The effects of social isolation and characteristics of the


Mack, M. (1980). The acquisition of stop systems: A cross-linguistic per...


Language and Stress and prominence in Finnish and English.


