Crosscurrents
in
Second Language Acquisition
and
Linguistic Theories

Edited by
Thom Huebner & Charles A. Ferguson
LANGUAGE ACQUISITION & LANGUAGE DISORDERS

EDITORS

Harald Clahsen  William Rutherford
Universität Düsseldorf  University of Southern California

EDITORIAL BOARD

Melissa Bowerman (Max Planck Institut für Psycholinguistik, Nijmegen)
Patricia Clancy (University of Southern California)
Werner Deutsch (Universität Braunschweig)
Kenji Hakuta (University of California at Santa Cruz)
Kenneth Hyltenstam (University of Stockholm)
Peter Jordens (Free University, Amsterdam)
Barry McLaughlin (University of California at Santa Cruz)
Jürgen Meisel (Universität Hamburg)
Anne Mills (University of Amsterdam)
Csaba Pleh (University of Budapest)
Michael Sharwood Smith (University of Utrecht)
Catherine Snow (Harvard University)
Jürgen Weissenborn (Max Planck Institut für Psycholinguistik, Nijmegen)
Lydia White (McGill University)
Helmut Zobl (Carleton University, Ottawa)

Volume 2

Thom Huebner and Charles A. Ferguson (eds)

Crosscurrents in Second Language Acquisition
and Linguistic Theories
CROSSCURRENTS IN SECOND LANGUAGE ACQUISITION AND LINGUISTIC THEORIES

edited by

THOM HUEBNER
San Jose State University

and

CHARLES A. FERGUSON
Stanford University

JOHN BENJAMINS PUBLISHING COMPANY
AMSTERDAM/PHILADELPHIA
1991
III. From Data to Model Building
Perception and production: the relevance of phonetic input to L2 phonological learning

James Emil Flege
University of Alabama at Birmingham

1. Introduction

Native-language (L1) acquisition research shows that children's initial attempts at realizing sounds often diverge considerably from the adult model. Despite this, children's misarticulations never seem to give rise to the perception of foreign accent. Adults know that the child's [wrebIt] means 'rabbit' because the segmental substitution in it is predictable on the basis of previous experience with “child-accented” English. The substitutions seen in nonnative speakers' production of sounds in a second language (L2) vary systematically according to their L1 background. For those who learn an L2 after the L1 has been thoroughly established, the substitutions do not derive from immaturity of the speech production mechanism or inability to fully control its functioning (Macken and Ferguson 1981), but rather from the perceptual misinterpretation of phonetic contrasts in the L2 and/or the inappropriate use of patterns of production used for L1 sounds in the production of L2 words.

Children succeed in realizing most sounds in their L1 in a recognizable, language-appropriate fashion within several years of the emergence of a working lexicon. For example, children acquiring Spanish as an L1 learn to implement /t/ as a voiceless unaspirated stop with a dental place of articulation. Children acquiring English, on the other hand, learn to implement /t/ as a voiceless aspirated stop with an alveolar place of articulation (Flege and Eefting 1986). There is nonetheless instrumental evidence that children's ability to realize L1 sounds continues to mature long after the sounds are said to have been “mastered” (e.g., Eguchi and Hirsh 1969; Flege, McCutcheon, and Smith 1987; Flege 1988c). Articulatory parameters in phonetic realizations rules are adjusted slowly, apparently so that acoustic
output will conform ever more closely to the phonetic norms specified in
central phonetic categories. Ability to reproduce the acoustic properties in
a class of phones (or “sound”, for short) in a reliable, rapid manner is the
hallmark of the human phonetic learning ability.

An important theoretical and practical question is whether the phonetic
learning ability remains intact beyond early childhood. There is evidence
to suggest that syntactic learning ability diminishes gradually (or is used
less) up to about the age of 12 years (Johnson and Newport 1989). More
heat than light has been shed in the continuing controversy concerning the
effect of age on the phonetic learning ability (see Flege 1990b for a review
and new data). Do adults become unable to establish phonetic categories?
Or unable to translate the sensory information encoded in phonetic
categories into the gestures needed to produce L2 sounds accurately? If so,
is it because they have passed a “sensitive period” for learning speech?

Much has been made of the similarity between L1 acquisition and L2
learning, but important differences do exist. Infants are able to discriminate
auditorily most, if not all, of the phonetic contrasts exploited by human lan-
guages. A large cognitive learning task awaits the child acquiring L1 pho-
nology. Young children only gradually discern from the phonetic input
available to them what are the phonemic and phonetic categories of their
L1. According to Aslin and Pisoni (1980), children’s perceptual processing
of L1 speech must be “attuned” in some instances to the phones in L1 to be
fully efficient for the processing of incoming speech.

Those learning an L2, on the other hand, have already established a
system of phonemes, and phonetic categories with which to implement
them. One might characterize child L1 acquisition as a “bottom-up” process
of learning, whereas adult L2 learning might be seen as largely a “top-
down” process (see Mack 1988, for a discussion). Like many others (e.g.,
Weinreich 1953; Wode 1977, 1978), we recognize the fundamental impor-
tance of L2 learners’ attempts to match or find correspondences between
phonetic elements in L1 and L2. Because of this, L2 learners’ perceptual proces-
sing of phones in the L2 is apt to be hindered by the earlier attunement
of auditory perception to phones in the L1.

Work in our laboratory has focussed on L2 speech production and per-
ception by adults who learned the L2 either as adults, or earlier in life as
children. When adults learn an L2, they are no longer actively engaged in
establishing phonetic and phonemic categories. Adults are generally more
analytic than children in their perceptual processing of the world around
them (e.g., Smith and Kemler Nelson 1984). They are also more likely than children to break down words into phoneme-sized units (see Walley 1990 for an excellent review). As a result, adults are more likely than children to identify L2 phones in terms of already-established L1 categories even though it may mean “forcing square pegs into round holes”. For example, native English adults tend to identify the initial stop in a Spanish word like todos as a realization of their English /t/ category even though the /t/ of Spanish is an unaspirated dental stop whereas English /t/ is an aspirated stop produced with an alveolar place of tongue constriction.

Most of our experiments to date have focussed on specific acoustic dimensions in phonetic segments, comparing the performance of extreme groups of L2 learners differentiated primarily by amount of previous L2 experience. The mean values observed for monolingual speakers of the L1 are used as benchmarks to chart L2 learners’ progress in modifying previously established patterns of L1 production as they gain experience in the L2. Mean values observed for native speakers of the target L2 are used to quantify the extent to which the phonetic norms of the L2 have been approximated (e.g., Flege 1980). This procedure rests, of course, on the hypothesis that the native speakers’ speech adequately represents the kind of L2 that has been encountered. (See Flege (1990c) for a discussion of the methodological importance of considering the various “Englishes” that may be learned.)

The results of the research summarized below support the belief that adults retain the phonetic learning ability that permitted them, as children, to learn to speak their L1 without accent even though certain factors may prevent them from making optimal use of their ability. Authentic production of a subset of the sounds in L2, referred to as “new” sounds, appears to be constrained only (or at least mostly) by the amount and quality of L2 phonetic input. As will be discussed below, the authentic production of another subset of the sounds of L2, referred to as “similar” sounds, may prove to be more difficult.

The results of research examining the ability of adult L2 learners to produce similar L2 sounds appear to support the prediction generated by the assumption that a sensitive period exists for speech learning (see Flege 1987b), but not for the reasons that have been traditionally assumed (i.e., general neurological maturation, lateralization for language functions, a diminished capacity for sensorimotor learning). Older children and adults often merely approximate the L2 phonetic norms for similar L2 sounds
(i.e., do not produce them authentically) because of equivalence classification. This mechanism seems to hinder or even to prevent the establishment of phonetic categories, and may cause L2 Learners to merge the acoustic characteristics of corresponding L1 and L2 sounds that have been identified with one another.

2. Speech Perception

One question we must consider is whether L2 learners are aware of systematic differences between sounds in L1 and L2. To what extent, for example, are Spanish/English bilinguals aware that the /t/ of Spanish (which is implemented as a voiceless unaspirated stop with short-lag VOT values)
differs from English /t/ (which is a voiceless aspirated stop realized with long-lag VOT values)? The existence of verifiable cross-language differences in speech perception (e.g., Flege and Hillenbrand 1987; Flege 1990c) sets the stage for this important question: Will equivalence classification cause L2 learners to “filter out” audible acoustic differences, thereby preventing them from establishing the accurate perceptual “targets” (or articulatorily specified phonetic goals) needed to guide speech learning?

The data shown in Figure 1 illustrates a well-known fact: if stops are produced differently in two languages, they will be perceived differently by native speakers of those languages. The identification functions, which are from Flege and Eefting (1986), plot the number of times each member of a VOT continuum ranging from /da/ to /ta/ was labelled “da” by subjects in a speech perception experiment. The subjects were adult and 8-year-old native speakers of English or Spanish. Of interest are differences between subject groups in the location of the “phoneme boundary”, that is, the points along the VOT continuum where responses shifted from predominantly “da” to “ta”. The native English adults and children (EA, EC) required about 15 msec more VOT than the native Spanish subjects (SA, SC) before giving predominantly voiceless stop responses (see also Williams, 1980).

This result shows that the basic auditory processes which underlie stop identification may be modified on the basis of experience hearing (and possibly producing) the stop consonants which occur in the L1. The main effects of Age and Language were significant, but the two-way interaction between the Language and Age factors was nonsignificant. Since much the same cross-language difference existed for adults and children, one might infer that the process of language-specific perceptual “attunement” (Aslin and Pisoni 1980) to the phonetic characteristics of L1 stops is largely complete by the age of 8 years. Additional developmental research is needed, however, since the age at which the attunement has been completed may be relevant to the issue of whether additional phonetic categories can be established for similar L2 sounds (see below).

2.1 Foreign accent perception

To help determine whether L2 learners are able to detect cross-language phonetic differences in speech production, Flege (1984) edited /t/ realizations from English words that had been produced by native speakers of
French and English. French /t/ (like the /t/ of Spanish) is implemented as a voiceless unaspirated stop with a dental place of articulation. Despite these differences from the long-lag alveolar /t/ of English, native speakers of English usually identify the [t]s which realize /t/ in French words with English /t/ realizations.

Pairs of English syllables were presented to English-speaking listeners, one containing an /t/ spoken by a native English speaker and the other an /t/ spoken by a native speaker of French. The native English listeners were able to identify which syllable in the pairs had a French-produced /t/ at significantly above-chance rates. This is especially impressive when one considers that the English /t/ realizations they judged to have been produced by a French speaker were probably not true “French” /t/s. The French talkers' realizations of English /t/ may well have moved in the general direction of English phonetic norms for, as will be discussed further, adult learners often approximate L2 phonetic norms. It may therefore have been even easier for the listeners to distinguish between /t/s spoken by English and French monolinguals.

In another experiment, Flege (1984) presented pairs of isolated release bursts that had been edited from English /t/ realizations spoken by native and French speakers of English. The native English listeners were able to choose the French-produced bursts at significantly above-chance rates. It appears that the small place of articulation difference between the dental /t/ of French and the alveolar /t/ of English was audible, and the listeners knew which was which. This finding suggests that adult listeners have a very detailed central representation of how the phones in their L1 “ought” to sound. It implies that L2 learners will be able to detect the divergence of L2 phones from the phonetic norms for L1 categories. But do they?

2.2 Foreign accent mimicry

One might argue that a paired-comparison task is not ecologically valid because listeners are generally more aware of meaning than sound. The rapid (and perhaps obligatory) assignment of phones to phonemic categories during speech processing makes us mercifully unaware of those aspects of speech production (phonetic implementation and realization) which determine the details of how particular phonetic segments have been articulated. It may be more difficult for L2 learners who are trying to understand what is being said in a conversation to detect subphonemic divergences of L2 phones from the norms for L1 categories.
An experiment by Flege and Hammond (1982) addressed this issue. Their experiment was based on the observation that native speakers of Spanish often realize /t/ in English with shorter VOT values than English native speakers. The speech question they raised was whether phonetically unsophisticated adults can detect the phonetic difference between the /t/s of Spanish and English. The study used a delayed mimicry task with 50 undergraduate students at the University of Florida who were familiar with Spanish-accented English. At the time of the study, the subjects were all enrolled in first-year-Spanish courses taught by native Spanish-speaking teachers. Some subjects were asked to read English sentences with what they considered to be a Spanish accent. Other subjects drawn from the same student population read the sentences normally (that is, without the instruction to speak with a feigned foreign accent).

The data in Table 1 indicate that the subjects were indeed able to speak with something that approximated a Spanish accent. This table contains information pertinent to seven English sounds that are known to be substituted by at least some Spanish speakers of English. More specifically, it presents the percentage of times each one of the sounds examined were transcribed as having been substituted by another segment in the expected Spanish-like way. If the subjects were tacitly aware that Spanish speakers devoice fricatives in word-final position, for example, we would expect

<table>
<thead>
<tr>
<th>Word</th>
<th>English Target</th>
<th>Spanish-accent Substitute</th>
<th>Frequency of Substitution</th>
</tr>
</thead>
<tbody>
<tr>
<td>nose, cheese, hose</td>
<td>/z/</td>
<td>/s/</td>
<td>141 (47%)</td>
</tr>
<tr>
<td>vice, veil, vase</td>
<td>/v/</td>
<td>/b/</td>
<td>129 (43%)</td>
</tr>
<tr>
<td>fig, pig, wig</td>
<td>/l/</td>
<td>/l/</td>
<td>127 (42%)</td>
</tr>
<tr>
<td>book, hook, crook</td>
<td>/U/</td>
<td>/o/</td>
<td>61 (20%)</td>
</tr>
<tr>
<td>shell, sheet, sheep</td>
<td>/j/</td>
<td>/i/</td>
<td>49 (16%)</td>
</tr>
<tr>
<td>bean, phone, bone</td>
<td>/n/</td>
<td>/n/</td>
<td>1 (0%)</td>
</tr>
<tr>
<td>tape, tube, toad</td>
<td>/t/</td>
<td>/d/</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

The frequency with which 50 native English-speaking students substituted a variant typical in Spanish-accented English for seven English target sounds when attempting to read English phrases with a Spanish accent. The absolute frequency of the substitutions are shown along with the percentages of times each target sound was affected. The data are taken from Flege and Hammond (1982).
them to show their knowledge by doing likewise. Indeed, the experimental subjects pronounced *cheese* (/ciz/) as [čis] and *nose* (/nozl/) as [nos]. They also produced /l/ as a stop (viz. [b]), /I/ and /U/ as tense vowels (viz. [i] and [u]), and the fricative /f/ as [č].

Somewhat surprisingly, the experimental subjects never produced /t/ as [d], even though Spanish /t/ is produced with short-lag VOT values. This raised the question of whether the experimental subjects were aware of **subsegmental** differences between native-produced and Spanish-accented English. More specifically, were they able to detect auditorily the fact that native speakers of Spanish tend to produce English /t/ with VOT values that are too short by English phonetic standards? If these differences can be (and

![Figure 2](image-url)

*Figure 2.* Mean voice onset time (VOT) of word-initial /t/s in English words said in the carrier sentence “The__is on the__” by 10 subjects in each of three groups, in msec. “Control” subjects read the sentences normally. Of the 50 experimental subjects, the 10 who produced the greatest number of Spanish-accent substitutions were designated “most knowledgeable” of Spanish-accented English, and those who produced the fewest were designated the “least knowledgeable” (see Table 1). Data are from Flege and Hammond (1982).
are) detected, could the subjects reproduce the relatively short VOT values that are characteristic of Spanish-accented English?

To answer these questions, VOT was measured in the /t/s spoken by the 10 subjects with the most knowledge of Spanish-accented English, that is, those who produced the largest number of segmental substitutions. It was also measured in stops spoken by the 10 subjects with the least knowledge of Spanish accent, and by 10 control subjects who read the sentences normally, that is, without mimicking a Spanish accent. Figure 2 shows that, as expected, the control group produced word-initial /t/s with the long-lag VOT values typical for English (ca. 80 ms). They did so both for stops in utterance-medial and utterance-final words.

The results for the experimental subjects revealed an awareness of cross-language phonetic differences. Both experimental groups produced /t/ with significantly shorter VOT values than the control subjects. Their VOT values ranged from about 40 msec to 60 ms, which is of course longer than the approximately 20-msec VOT values typical for English /d/ (or Spanish /l/, for that matter). The experimental subjects came close to the “compromise” VOT values we have come to expect in Spanish-accented English (see below). The experimental subjects who produced the largest number of segmental substitutions (“most knowledgeable”) decreased VOT more than those who produced the fewest substitutions (“least knowledgeable”).

If the number of segmental substitutions reflected amount of exposure to Spanish-accented English, one might infer that L2 learners become increasing aware of subsegmental phonetic differences as they gain experience in L2. Since the native speakers of English were able to detect and to reproduce the VOT difference between Spanish-accented and standard realizations of English /t/, it should be even easier for English learners of Spanish to note acoustic differences between the /t/ realizations of Spanish and English monolinguals, and vice versa. If so, difficulty in producing stops with correct VOT values should probably not be attributed to an auditory limitation per se.

2.3 Gauging degree of foreign accent

A study reviewed earlier (Flege, 1984) showed that native speakers can detect a foreign accent. They might do so by noting the replacement of one already familiar L1 sound for another (e.g., [bIt] for beat), or they might hear the intrusion of an L2 sound not found in the L1 (e.g., a German uvu-
lar /r/ in the English word bring). Foreign accent detection might also occur in the absence of overt sound substitutions as the result of perceived divergences of an input phone from an internalized phonetic norm. A distinction we find useful is the one between "adequacy", which refers to whether a sound is perceived as intended, and "acceptability", which refers to whether a correctly identified sound appears to be distorted. For example, a native Spanish speaker's somewhat [l]-like rendition of the vowel in English beat (/bit/) might differ sufficiently from the English norm for /l/ that it is judged to be foreign-sounding.

Flege (1988b) carried out a foreign accent study to help determine whether foreign accent perception can indeed be based on divergences from L1 phonetic norms in addition to overt substitutions, and whether nonnatives can perceive foreign accent in the L2 even if they themselves speak it with a foreign accent. The study examined the perception of degree of foreign accent in English sentences by three groups of listeners: native speakers of English, Chinese subjects who had lived in the U.S. for about five years, and Chinese subjects who had lived in the U.S. for only about one year on the average. The listeners estimated the degree of foreign accent by moving the lever on a response box between scale endpoints marked "strong foreign accent" and "no foreign accent". Variation in the lever's position lead to digitized values ranging from 1 (for the strongest accent) to 256 (for the most authentic possible accent). The English sentences evaluated were spoken by a group of native English talkers and by groups of Chinese talkers differentiated according to age of L2 learning and/or amount of previous English-language experience.

No difference between the experienced and inexperienced Chinese listeners was expected if foreign accent judgments are based on perceived sound substitutions alone. This prediction was based on the assumption that it would be equally easy for the experienced and inexperienced Chinese listeners to detect whatever Chinese-derived sound substitutions existed in the English sentences they were rating. If, on the other hand, foreign accent perception is influenced by subsegmental (or nonsegmental) divergences from phonetic norms, better performance was expected from the experienced than the inexperienced Chinese listeners. This prediction was based on the assumption that as the Chinese listeners gained experience with English, they would develop central representations for how English categories ought to sound and this, in turn, would make it easier for them to note divergences from English phonetic norms.
PERCEPTION AND PRODUCTION

Rating of Native and Non-native Speakers by 3 Listener Groups

Figure 3. Mean foreign accent rating accorded English sentences by three listener groups using a scale ranging from "strong foreign accent (1) to "no foreign accent" (256). The listeners were English monolinguals (ENG) and Taiwanese speakers of English who had lived in the U.S. for about one or five years (GrA, GrB). The talkers who had produced the sentences were native Mandarin speakers who had lived for about one year in the U.S. (M), Taiwanese subjects who had lived for one or five years in the U.S. (A, B), adults from Taiwan who learned English as children (C), and native speakers of English (E). The data are from Flege (1988b).

Figure 3 shows how the three listener groups in the Flege (1988b) study rated the English sentences produced by the five groups of talkers. The results show that L2 learners develop a notion of how English sentences ought to sound. The results imply, further, that L2 learners develop central representations for English phonetic categories. The ratings accorded to the talkers by the various groups of listeners were largely as expected. Sentences spoken by native speakers of English (Group E) and by Chinese subjects who had learned English at an average age of 7.6 years (Group C, for "child learners") received higher ratings than sentences produced by talkers in three other groups: native Mandarin-speaking subjects who had lived in the U.S. for about one year (Group M), and native Taiwanese subjects who had lived in the U.S. for about one year (Group A) or five years (Group B).
Figure 3 also shows an important difference between the three listener groups. Native English-speaking listeners (ENG) differentiated native and nonnative talkers (Group E versus Group M) to a significantly greater extent than the experienced Chinese listeners (GrB) who, in turn, differentiated native from nonnative talkers to a significantly greater extent than the inexperienced Chinese listeners (GrA). Assuming that subjects in both listener groups responded in the same way to segmental substitutions, these results suggest that the experienced Chinese listeners had a better notion of how the English sentences ought to have sounded than the inexperienced Chinese listeners. They may have done so by virtue of having established phonetic norms for sounds in the English sentences.

One finding in Figure 3 bears on the effect of age of learning on L2 speech production. There seems to be widespread belief that no foreign accent in an L2 will be heard if it is learned in childhood. The Chinese talkers in Group C received significantly lower scores than the native English talkers even though they had begun learning English at an average age of 7.6 years. The difference between Groups C and E indicates that the sentences of the Chinese “child learners” (Group C) were foreign-accented, which means that exposure to an L2 during childhood does not guarantee accent-free pronunciation of the L2. It is worth noting that most of the subjects in Group C attended elementary and high schools in Alabama, where most teachers are monolingual speakers of English. Since these subjects reported using Chinese at home, their foreign accents probably did not derive from foreign-accented input.

One aspect of the data in Figure 3 bears on the effect of amount of L2 experience on authenticity of L2 speech production. The Taiwanese subjects in Groups A and B all had begun learning English as adults. The Group A subjects had lived for one year in the U.S. whereas the Group B subjects had lived for five years in the U.S. The foreign accent scores of the two groups did not differ significantly, suggesting that the amount of L2 experience may have little influence on the L2 pronunciation of adult learners. Perhaps the difference in amount of meaningful English-language input that subjects in Group A and Group B had been exposed to was less than the difference implied by the length of residence differences. The Group B subjects spoke Chinese frequently on a daily basis, which may have minimized the amount of English-language input they received.

Despite the apparent lack of an effect of LOR on global foreign accent, it is unlikely that subjects differing substantially according to
amount of L2 experience will produce all English vowels and consonants with an equal degree of authenticity. As will be discussed below, certain L2 sounds are learned more thoroughly — and perhaps more rapidly — than others. An acoustic study of vowel production recently completed in our laboratory (Wang 1988) is worth mentioning in this regard. A relatively experienced group of Mandarin subjects (about five years in the U.S.) was better able to produce and perceive a contrast between /i/-/I/ than a group of less experienced subjects (about 1 year in the U.S).

Finally, the results in Figure 3 show a dichotomy between speech production and perception. The subjects in the inexperienced Chinese listener group (GrA) participated as talkers in Group A, and subjects in the experienced Chinese listener group (GrB) were drawn from talker Group B. Even though the Group A and Group B talkers spoke with equally strong foreign accents (as assessed by native English-speaking listeners), the GrB listeners were nevertheless better able than the GrA listeners to differentiate native from nonnative speakers of English. This shows that the ability to gauge degree of accent in L2 sentences may increase more rapidly (or to a greater extent) than ability to produce L2 sentences authentically.

2.4 Cross-language difference in speech perception

The results presented so far suggest that L2 learners can detect acoustic differences between sounds in L1 and L2, and may establish representations for L2 sounds. But do they use the same central phonetic category representations as natives to produce and perceive sounds in the L2? And do they process sounds in L2 just like native speakers?

A number of previous studies have shown perceptual differences between natives and nonnatives (e.g., Jamieson and Morosan 1986; Underbakke, Polka and Strange 1988; Flege 1988c). To illustrate the kind of differences that may exist, we will consider a study by Flege and Hillenbrand (1985) which examined the perception of the voicing feature in word-final fricatives. The results of this study suggested that L2 learners may be able to identify L2 sounds correctly, but it raised the issue of whether they differed from native speakers in how they perceptually process those sounds.

For native speakers of English, the distinction between /s/ and /z/ in word-final position is cued perceptually by a number of acoustic dimensions including vowel and fricative duration. For example, the vowel in peas (/piz/) is longer than that in peace (/pis/) and, conversely, the fricative noise
Effect of vowel duration on voicing judgments (data averaged over 5 fricative durations)

![Graph showing the mean frequency of /z/ judgments given in response to the members of the 25-member synthetic continuum ranging from /pis/ to /peas/, with five 50-msec increases in vowel duration. The data are from Flege and Hillenbrand (1985).]

Effect of Fricative Dur. on voicing judgments (data averaged over 5 vowel durations)

![Graph showing the mean frequency of /z/ judgments given in response to the members of the 25-member synthetic continuum ranging from /pis/ to /peas/, with five 50-msec increases in fricative duration. The data are from Flege and Hillenbrand (1985).]
of /z/ is shorter than that of /s/. Flege and Hillenbrand (1985) examined the response of native and nonnative speakers of English to co-varying changes in vowel and fricative duration in a 25-member continuum of synthesized CVCs ranging from *peas* to *peace*. Both vowel and fricative duration parameters were varied orthogonally in five 50-msec steps. Two groups of native English speakers differing in dialect (Midwestern versus Southern) participated. In addition, experienced and inexperienced native speakers of French, Swedish and Finnish took part. The French subjects served as a kind of control group, for the word-final /s-/l/z/ contrast in French is similar to that of English whereas both Swedish and Finnish have no such contrast owing to the absence of /z/.

Figure 4 (top) shows the effect of the vowel duration variations. The mean values plotted here are based on the responses of two subgroups for each language, since neither the effect of dialect (for the native speakers of English) nor the effect of English-language experience (for the nonnatives) had an important effect on the voicing judgments. The percentage of /l/z/ judgments increased as vowel duration increased for the subjects in all four groups. As expected, the effect was somewhat greater for the English and French subjects than for the Swedish and Finnish subjects. Interestingly, the French subjects made significantly less use of vowel duration than the English subjects, probably because the voicing feature of a word-final fricative affects the duration of a preceding vowel less in French than English (at least in pre-pausal position).

Figure 4 (bottom) shows that fricative duration had less effect on fricative voicing judgments than did the equal variations in vowel duration. For the English and French subjects, /l/z/ judgments decreased significantly as fricative duration increased. Fricative duration did not have a significant effect on the voicing judgments of the Swedes or Finns, however.

Although the subjects in all four groups were able to divide the continuum into two categories based on changes in vowel duration, this does not mean necessarily that the vowel duration changes actually resulted in perceptual differences. The Finns and Swedes, being familiar with the phonemic use of vowel duration in their L1, might simply have devised the strategy of circling /l/z/ on the answer sheet when they heard a relatively long vowel. The reversals noted for some subjects (that is, the use of a /l/z/ label for stimuli with short rather than long vowels) is consistent with a "conscious strategy" interpretation. The results strongly suggest that neither the Swedes nor the Finns used fricative duration as a cue to the /s/-/l/z/ contrast.
in English, perhaps because the vowel duration differences were more obvious than the fricative differences owing to relatively greater intensity of vowels than fricatives or the nature of phonemic length distinctions in the L1. Whatever the explanation, it is clear that the Swedish and Finnish subjects did not — like the native speakers of English — integrate the co-varying vowel and fricative duration cues to arrive at /s/ and /z/ percepts. This, in turn, suggests that their central representations for those categories differed from the native English speakers'.

3. Speech Production

Unless nonnatives develop L2 phonetic categories that are the same (or similar) as native speakers', they may be unable to produce L2 phones authentically, at least insofar as language-specific aspects of speech production are concerned. This hypothesis rests on the assumption that, in speech learning, perceptual representations derived from experience with phones in the ambient language guide development of the motor programs needed to implement phonemic categories. For example, if a correct perceptual representation is needed for authentic production, then one would not expect to see a difference in the duration of /s/ and /z/ frication for Finnish and Swedish learners of English.

The notion that "perception leads production" is undoubtedly too simplistic to explain the complex interaction of speech production and perception that unfolds during speech learning. However, like others (e.g., Strange and Broen 1981), I nevertheless find it a useful heuristic. The notion implies that sensory-based perceptual information and information pertaining to phonetic implementation (and perhaps realization) are encoded in distinct, often nonisomorphic, central representations. It implies further that at a certain stage of L2 learning, a nonnative might perceive an L2 sound just like native speakers, but not be able to produce it authentically. Just as it takes children some time to master certain L1 sounds (see Wode 1978; Macken and Ferguson 1981), it may take L2 learners some time to learn how to produce sounds according to the plan encoded in phonetic representations.

We would not expect to see the opposite pattern, that is, nonnatives who fail to perceive an L2 sound like native speakers but always produce it authentically. One well-controlled study seemed to provide counterevi-
dence to this claim, however. Sheldon and Strange (1982; see also Brière 1966) found that the frequency with which English /r/s and /l/s spoken by Japanese subjects were correctly identified by native English listeners exceeded the rate at which the Japanese subjects correctly identified /r/s and /l/s spoken by native English talkers. The divergence may have derived from the Japanese subjects’ previous formal training in English (see p. 254). That is, their success in producing English liquids may have derived from the use of an articulatory strategy such as “To say /r/, place the tongue....” or “To say /l/, combine the features of the Japanese X and Y sounds”. If so, then their production of English might not have appeared to be better than their perception of the same segments had their production of liquids in connected speech been examined (see e.g., Dickerson 1974) or had their relatively “unmonitored” production of English been assessed (see Flege and Hillenbrand 1984).

The speech perception research just reviewed suggests that L2 learners can detect phonetic differences between sounds in L1 and L2 and may develop perceptual norms for L2 sounds. They may not, however, make use of the same perceptual cues as native speakers when identifying L2 sounds (assuming they can do so reliably). How, then, will they produce sounds in L2?

3.1 New versus similar sounds

It is beyond the scope of this chapter to review the numerous L2 speech production studies that have been carried out in the past decade (for a review of studies up to 1984, see Flege 1988a). This section will therefore focus on just a few studies examining the production of vowels and the voicing feature in word-initial stop consonants. The results of these studies provide preliminary evidence that “new” sounds may be learned better than “similar” sounds. For operational procedures that might be used to define sounds in an L2 as new or similar, as well as a more detailed elaboration of the model sketched here, see Flege (1990a).

The distinction between new and similar sounds is by no means a novel one. For example, Delattre (1964; 1969) noted that some sounds in an L2 differ “radically” from any sound in the L1 and should be regarded as new from the standpoint of the L2 learner. Wode (1978: 114) noted that a major difference between child and adult learners of an L2 is “the state of development” of their phonological systems. In his view, both children and
adults match phonetic elements of the L2 to their L1 “grid”. As the L2 is processed, the acoustic input is “scanned” and phones falling within some “crucial similarity range” are judged to be equivalent to an element of L1, and therefore substituted by it. Other phones falling outside the crucial (but undefined) range are judged to be “non-equivalent”. Wode (1978: 114) says that such a phone will undergo “other developments” than simple substitution. (On this, more later.)

A valid and reliable method has yet to be developed for determining which sounds of an L2 will be treated by L2 learners as new and which as similar. The question that arises is: When does an L1 versus L2 acoustic difference make a phonetic difference? Flege (1987a; see also Flege and Hillenbrand 1984; Flege 1990a) used a simple test. He suggested that, as a first approximation, an L2 sound should be considered new if it is represented by a phonetic symbol not used to represent any sound in the L1. This approach makes indirect use of the expert phonetic classifications of linguists and phoneticians who have worked with the languages under consideration.

The problem with this “phonetic symbol” method is that the experts don’t always agree. For example, the distinction between English /i/ and /ɪ/ is sometimes represented as just shown, and sometimes as /i:/ versus /ɪ/. One transcription emphasizes the vowel quality difference between this pair of high front unrounded tense and lax vowels whereas the other emphasizes their duration difference. The use of different symbols can lead to different predictions concerning how Spanish learners of English, for example, would produce English vowels. According to the first symbolization, the English lax vowel would be regarded as new for Spanish learners of English, whereas it would be regarded as similar according to the second symbolization. Even if phoneticians always agreed in their choice of phonetic symbols, their judgments might not be predictive of how untrained L2 learners relate the sounds of two languages.

Other methods for classifying L2 sounds as new or similar are possible. Bohn and Flege (1989) suggested that an L2 vowel phoneme might be identified as new on the basis of acoustic analysis. New vowels would be those whose realizations occupy a portion of an acoustic phonetic vowel space (e.g., F₂ versus F₁) that is unoccupied by the realizations of any L1 vowel. Token-to-token variability might be used as a criterion for determining whether a sound in L2 is new or similar. An L2 sound without an equivalent in the L1 might be produced more variably than one judged to have an L1 equivalent. This is because similar sounds tend to be substituted by a single
L1 category by even inexperienced L2 learners whereas new sounds may be substituted by a range of variants, at least in early stages of L2 learning. Ultimately, however, laboratory tests will be needed to determine whether sounds in L2 should be classified as new or similar. The status of an L2 sound vis-à-vis sounds in the L1 may influence on-line speech processing measurably. New sounds might require a longer time to identify than similar sounds in a speeded classification test; similar sounds may take longer than L2 sounds that are identical to sounds in the L1. In some instances it may be necessary to use a labor-intensive process of phonetic distance scaling to determine the relationships between sounds in L1 and L2.

A hypothesis that has guided a number of experiments in our laboratory is that adult L2 learners will eventually be more successful in producing new than similar sounds because they are able to establish phonetic categories for new phones (e.g., Flege 1987a, 1988a). The basis for the hypothesized difference is that similar sounds are classified as equivalent to L1 sounds and therefore substituted by L1 sounds (at least initially), whereas separate phonetic categories may be established for new sounds because they evade equivalence classification.

One could hypothesize just the opposite, of course. Based on a short-term laboratory experiment, Briere (1966: 795) concluded that L2 sounds "which are close equivalents to (L1) sounds, whether phonemic or allophonic" will be easier to learn than L2 sounds "without such equivalents".

3.2 Vowel production

A study by Flege (1987a) tested the hypothesis that new sounds will be produced more authentically than similar sounds. The study examined Americans’ production of French vowels. The high front rounded /y/ of French was classified as new, and French /u/ was classified as similar. This was done because the same phonetic symbol is used to represent the /u/ of French and English even though realizations of these vowels differ acoustically and auditorily (Flege 1984; Flege and Hillenbrand 1984). The hypothesis about new versus similar sounds predicted that native English learners of French would produce French /y/ more authentically than French /u/.

Figure 5 (top) presents mean F2 frequency data obtained by Flege (1987a) for French /y/ and /u/ as spoken by subjects in native English groups
Figure 5. (top) The mean F2 frequency in Hz of the vowel in French tu /tu/ (unfilled bars) and French tous /tus/ (filled bars) spoken by native speakers of English who had lived in France from as little as 9 months to as much as 12 years (Groups EF-1 to EF-3), and by native speakers of French who had lived in the U.S. for 12 years (FE-3). The values for French monolinguals are also shown. (bottom) The mean voice onset time (VOT, in msec) in the tu/ of English two (unfilled bars) and French tous (filled bars) spoken by bilingual subjects (EF1-3, FE-3). Mean VOT values for French and English monolinguals are shown by horizontal lines. Most means in both panels are based on 70 measurements; the brackets enclose +/- one standard deviation. The data are from Flege (1987a).
who had lived in a French-speaking environment from as short a time as 9 months to as long as 12 years (EF-1 to EF-3), French speakers of English who had lived in Chicago for 12 years (FE-3), and French monolinguals. The results for French /y/ are consistent with the prediction that adults can learn to produce new L2 vowels authentically. The least experienced native English speakers of French (EF-1) produced French /y/ with significantly lower Fz values than the French monolinguals. Native English speakers who were more experienced in French (Groups EF-2 and EF-3), on the other hand, did not differ significantly from the French monolinguals.

The results for French /u/ are consistent with the hypothesis (Flege and Hillenbrand, 1984) that equivalence classification will place an upper limit on how authentically similar sounds can be produced by adult L2 learners. The results for the three native English groups showed that Fz values in French /u/ became increasingly French-like (that is, decreased) as a function of L2 experience. However, even the most experienced native English group (EF-3) produced /u/ with significantly higher (and therefore English-like) Fz values than the French monolinguals. The native English subjects in EF-3 had used French as their primary language for over a decade, so this finding supported the claim that even highly experienced L2 learners will not produce similar L2 sounds authentically.

3.3 Stop consonant production

Following the pioneering work of Caramazza et al. (1973), we have examined the production of /p,t,k/ in English L2. As mentioned earlier, our original assumption was that L2 learners of all ages equate the long-lag voiceless stops of English with the short-lag voiceless stops of their L1 at a categorical level despite the existence of audible acoustic differences. Flege (1987a) examined the production of French and English /t/ by adult L2 learners. Figure 5 (bottom) provides additional confirmation of the hypothesis that an upper limit exists on how authentic the production of a similar L2 sound can be. This figure plots the mean VOT values measured for /t/ in comparable French and English words (French tous /tu/ and English two /tu/).

As shown by the horizontal lines in Figure 5 (bottom), /t/ was produced with substantially greater VOT value by English monolinguals than French monolinguals. Even highly experienced adult L2 learners did not produce /t/ authentically in L2. The subjects in FE-3 (native French speakers who
had lived in Chicago for 12 years) produced English /t/ with significantly shorter VOT values than English monolinguals. The native English subjects in EF-2 had lived for over a year in France. They produced French /t/ with VOT values that were significantly longer than the French monolinguals' /t/.

Contrary to expectation, the most experienced English speakers of French in Group EF-3 did not produce French /t/ with significantly longer (English-like) VOT values than the French monolinguals. The lack of a difference might be attributed to the mean value established as the French phonetic norm, which was suspiciously high. (The mean value for the French monolingual group was high because of the values for two subjects who seem to have accommodated their production of French /t/ to that of the experimenter (JEF), who elicited data using English-accented French.)

Another possible explanation for the “native-like” performance of the subjects in Group EF-3 for L2 /t/ is that equivalence classification does create the hypothesized barrier to the authentic production of similar sounds for most L2 learners but that, by chance, Group EF-3 included a disproportionate number of individuals who were especially gifted for L2 learning. (Perhaps a special talent for speech learning contributed to their decision to live in Paris.) Without a better understanding of the range of individual differences possible, the basis of the unexpected finding for Group EF-3 is uncertain. It is worth noting, however, that two recent studies have shown that a small percentage of adult Dutch (Flege and Eefting 1987b) and Spanish (Flege and Bohn 1989b) learners of English managed to produce English /t/ with values that fall within the range observed for English native speakers.

3.4 Age of learning

For the most part, however, adult L2 learners are unable to produce similar L2 sounds authentically. A VOT experiment by Flege and Eefting (1987a) tested the hypothesis that similar sounds can be learned authentically if L2 learning begins by about the age of 5-6 years. It was predicted that the VOT values in English /p,t,k/ produced by Spanish subjects who learned English L2 by that age would not differ significantly from English monolinguals' mean values.

The Flege and Eefting (1987a) study examined English stops produced by two groups of Puerto Rican adults who had learned English as children.
Figure 6. Mean VOT (in msec) in voiceless stops spoken by adults (top) and children (bottom). The values produced in English /p,t,k/ by Spanish/English bilinguals (ECB, LCB, BC) are compared to values for English monolinguals (“English”), and also to /p,t,k/ in Spanish words produced by Spanish monolinguals (“Spanish”). See text for a description of the three bilingual groups. The data are from Flege and Eefting (1987a).
Although both groups consisted of Early L2 learners, the subjects in one group had learned English at a slightly earlier age than those in the other group. The “earlier childhood bilinguals” (ECB) had spent most of their early childhood in New York and had attended an English-speaking elementary school there. The “later child bilinguals” (LCB) had little or no prior exposure to English before enrolling in a bilingual elementary school in Puerto Rico where English was one of the languages of instruction.

Figure 6 (top) shows that the subjects in ECB and LCB produced English /p,t,k/ with substantially longer VOT values than were observed in the production of utterance-initial Spanish stops by Spanish monolinguals. However, contrary to hypothesis, they produced English /p,t,k/ with significantly shorter VOT values than English monolinguals.

A similar finding was obtained for Puerto Rican children (BC) who had learned English as an L2. Figure 6 (bottom) presents the mean VOT values measured in stops spoken by monolingual English and Spanish children, and in the English stops produced by a group of native Spanish 8-year-olds who had been attending an English/Spanish bilingual school in Puerto Rico for about three years (designated Group BC, for “bilingual children”). The children in BC spoke English with relatively little foreign accent, in my opinion, but they nevertheless produced English /p,t,k/ with significantly shorter VOT values than the monolingual English children. These segmental acoustic findings agree with the results of the foreign accent experiment described earlier (Flege 1988b) and a VOT study by Caramazza et al. (1973) in showing that learning L2 in early childhood does not guarantee authentic production of the L2.

The data obtained for the Puerto Rican subjects might be regarded as a disproof of the age of learning hypothesis. If so, then one might conclude that an important aspect of the human phonetic learning ability is lost by about the age of 5-6 years. Perhaps not even young children are able to establish separate phonetic categories for English /p,t,k/ if voiceless stops are implemented as voiceless unaspirated stops in the L1. In my opinion, however, a more likely explanation of the Puerto Rico data exists. The Puerto Rican subjects examined in the Flege and Eefting (1987a) study may have received accented L2 input.

3.5 The input hypothesis

The Puerto Rican subjects in the Flege and Eefting (1987a) study lived in Mayaguez, a city where Spanish is clearly the predominant language. Eng-
lish seems to be used in Mayaguez mostly in official contexts. It is seldom heard on radio or TV. Movies in English are dubbed into Spanish at local theaters. I speculate that much — maybe even most — of the English heard by the Puerto Rican subjects in groups ECB, LCB, and BC was Spanish-accented. If so, then they may have produced English /p,t,k/ with nonauthentic VOT values because that is what they **heard**.

Recall that /p,t,k/ are typically realized in Spanish-accented English with VOT values that are intermediate to the phonetic norms of Spanish and English. If the native Spanish subjects examined by Flege and Eefting (1987a) established phonetic categories for the long-lag stops of English, it is likely that the VOT values specified in these categories were based on all the stops they had heard. The Puerto Rican subjects probably heard English stops realized with a wider range of VOT values than English monolinguals, and perhaps with a preponderance of values intermediate to those normally found in stops spoken by Spanish and English monolinguals. If so, then their English /t/ category would be expected to specify a shorter VOT value than English monolinguals’ (and a longer VOT value than in Spanish monolinguals’).

If the Puerto Rican subjects realized stops with the intermediate VOT values hypothesized to have been encoded centrally, we would expect them to **realize** English /p,t,k/ with shorter VOT values than native English speakers. The input hypothesis implies that the Puerto Rican subjects would not **have** differed from English monolinguals had they entered into frequent contact with native English speakers at the time during childhood when they established phonetic categories for English /p,t,k/. The native French subjects examined by Caramazza et al. (1973) may also have produced English stops with VOT values that were too short by English phonetic standard because of accented input (Yeni-Komshian, personal communication, 1989).

Support for the “accented L2 input hypothesis” was provided by a recent study examining the performance of Early L2 learners in Texas who received unaccented L2 input in childhood. Flege (1990c) examined the English and Spanish spoken by native speakers of Spanish in Texas. Like the Puerto Rican subjects in the study discussed earlier, the Early L2 learners in the Texas study were first massively exposed to English when they began elementary school at the age of 5-6 years. However, unlike the Puerto Rican subjects, the Texas subjects were exposed to **native**-produced English from the start. A majority of their classmates or their teachers in
grades 1-3 were native English speakers. The Texas subjects heard English spoken on the street, and on radio and TV.

VOT was measured for Spanish and English monolinguals, and in the L1 and L2 speech of the Early L2 learners and Late L2 learners from Texas who began learning English L2 at an average age of about 18 years. As in previous VOT studies, the Late L2 learners in the Flege (1990c) study produced /t/ with significantly longer VOT values in English than Spanish, but their English /t/ nevertheless had significantly shorter VOT values than English monolinguals'. As predicted by the input hypothesis, the Early L2 learners, on the other hand, did not differ from the English monolinguals. The Early L2 learners also produced Spanish /t/ with VOT values that did not differ significantly from those of Spanish monolinguals. Insofar as their production of /t/ in Spanish and English was concerned, the Early L2 learners were perfectly bilingual.

Additional support for the "accented phonetic input" interpretation was provided by an experiment examining stop perception. The purpose of this experiment was to provide insight into why the Puerto Rican subjects produced English /p,t,k/ with shorter VOT than the English monolinguals in the Flege and Eefting (1987a) study. The Spanish subjects identified the

Figure 7. The mean /t/-/d/ phoneme boundaries obtained in a forced-choice identification experiment examining listeners in three child and four adults groups made up either of monolinguals ("Spanish", "English") or Spanish/English bilinguals (BC, LKCB, ECB), in msec. See text for a description of the three bilinguals groups. The data from Flege and Eefting (1987a).
members of a VOT continuum ranging from /da/ to /ta/. As shown in Figure 7, the mean phoneme boundaries obtained for the adults in groups ECB and LCB (see above) occurred at significantly shorter VOT values than those of the native English adults. The fact that the ECB and LCB subjects were willing to apply the /t/ label to stimuli with shorter VOT values than the English monolinguals suggests that they processed long-lag stops differently than the English monolinguals, just as one would expect if their phonetic representations for English /p,t,k/ were based on Spanish-accented input.2

3.6 Imitation of stop consonants

Further support for the "accented phonetic input" interpretation of the Puerto Rico data came from an imitation study involving the same Puerto Rican subjects whose spontaneous English production was examined by Flege and Eefting (1988). The subjects in five groups mimicked the members of a VOT continuum ranging from /da/ to /ta/. (This was the same continuum whose members had been identified in the identification experiment just described.) The VOT values of the stimuli ranged from -60 msec to 90 ms, which spanned the three modal VOT categories used to implement phonologically voiced and voiceless stops in human languages (i.e., lead, short-lag, and long-lag). If the bilingual subjects in groups ECB, LCB, and BC had established separate phonetic categories for English /p,t,k/, it was predicted that they would imitate the long-lag stimuli with long-lag VOT values but, if not, they were predicted to imitate both short-lag and long-lag VOT stimuli with short-lag VOT values. This assumed, of course, that the subjects would identify stimuli to be imitated in terms of existing phonetic categories.

The results obtained by Flege and Eefting (1988) supported the assumption concerning covert identification. The subjects did not reproduce accurately the VOT values in the stimuli: there were substantial nonlinearities in the relationship between the stimulus VOT values and the VOT values measured in the subjects' vocal responses. At some point along the continuum, the VOT values in the subjects' responses increased by a much greater than average increment. The location of these nonlinearities coincided with the phoneme boundaries obtained in the earlier identification experiment (Flege and Eefting 1987a).
The frequency of all of the VOT values measured in the subjects' vocal responses were plotted in histograms, one for each group. As expected, there was a bimodal distribution of values in the monolingual Spanish subjects' histogram, with peaks in the lead and short-lag VOT ranges. For example, the monolingual Spanish children showed concentrations of values in the lead and short-lag ranges (more specifically, peaks in the frequency distribution centered at -114 msec and 14 ms). These subjects generally produced a narrow range of lead VOT values in imitating the six stimuli with lead VOT values, and produced a narrow range of short-lag VOT values when imitating the stimuli with both short-lag and long-lag VOT values. The monolingual English subjects also showed a bimodal distribution, with concentrations of VOT values in the short-lag and long-lag ranges. The native English children, for example, had a concentration of values in the short-lag range (i.e., a peak centered at 14 ms) and another in the long-lag range (at 80 ms). Unlike the monolingual Spanish children, the monolingual English children generally imitated the lead stimuli with short-lag VOT values.

The VOT values measured in the bilingual subjects' imitation responses supported the hypothesis that the subjects in groups EC, LCB, and ECB had established separate phonetic categories for English /p,t,k/. The histograms prepared for the bilingual subjects showed a trimodal distribution of VOT responses that was a composite of the patterns evident in the Spanish and the English monolinguals' histograms. The bilinguals in all three groups showed concentrations of VOT values in the lead, short-lag, and long-lag ranges.

It seems reasonable to think that the monolingual Spanish and English subjects perceived two phonological categories in the VOT continuum (viz., /d/ and /t/ phonemes) although they divided the continuum at different places, that is, had different phoneme boundaries. It is likely that the bilinguals also divided the VOT continuum into two phonological categories but, by hypothesis, they differed from the monolinguals in an important way. Unlike the monolinguals, the bilinguals were able to implement /t/ in two ways using different phonetic categories, one for Spanish short-lag /h/ and one for English long-lag /t/. As a result, they were able to reproduce stimuli with VOT values ranging from 0 msec to 90 msec more accurately than monolingual subjects by responding differently to [t] and [tʰ] phones in the 2.5-sec intervals between the successive VOT stimuli being imitated.
3.7 Effects of L2 learning on L1 speech production

One other type of evidence was provided in support of the hypothesis that nonauthentic phonetic input, not inability to establish new phonetic categories, caused the early L2 learners in the Flege and Eefting (1987a) study to differ from native speakers of English. This evidence is more subtle than the evidence presented thus far, and so requires some introduction. The evidence has to do with differing effects of L2 learning on the production of stops in L1. It appears that VOT values in L1 stops may decrease or increase depending on age of L2 learning.

By hypothesis, the central representations for /p,t,k/ of adult French and Spanish learners must serve as a basis for the identification and production of stops in L1 and also for similar stops in English L2. As the result of previous experience with short-lag voiceless stops in their L1 and long-lag voiceless stops in English, the representations for /p,t,k/ of native speakers of French and Spanish who learn English will change. The process is likely to occur slowly, given the many years it takes children acquiring English as an L1 to fully establish productive and perceptual control of the VOT dimension in English stops (see e.g., Macken and Barton 1980, and references in Flege and Eefting 1986). In time, French/Spanish learners' perceptual representations for /p,t,k/ may specify a VOT value that reflects all of their previous L1 and L2 phonetic input, thereby merging the acoustic phonetic properties of similar L1 (French or Spanish) and L2 (English) stops. One would expect such representations to specify VOT values that are intermediate to those of L1 and L2 monolinguals.

If the view of L2 speech learning represented here is correct, then some additional mechanism is needed to account for the fact that adult French or Spanish learners of English typically produce /p,t,k/ in English with significantly longer VOT values than in the L1. Flege and Eefting (1988) suggested that they do so by using different realization rules to output /p,t,k/ categories in L1 and L2. Just as speakers who control two dialects of English may realize vowels differently as a function of social context (e.g., Labov 1983), adult L2 learners may produce /p,t,k/ differently in L1 and L2 by means of relatively small adjustments in production parameters, not by selecting different phonetic categories with which to implement the phonemes /p,t,k/ (see also Flege and Bohn 1989b, for a recent discussion).
Figure 8. Illustration of the differing effect of L2 learning on L1 stop production for Early and Late L2 learners. (top) Mean VOT (in msec) in L1 /t/s spoken by monolinguals (“French”, “English”), experienced French speakers of English (B1, left side), and three native English groups who had all learned French as adults but were differentiated by increasing L2 experience (B1-B3, right side). (bottom) Mean VOT in Spanish /t/ spoken by child and adult monolingual speakers of Spanish (“Spanish”), native Spanish children who spoke English (BC), and two groups of native Spanish adults who had learned English in childhood (ECB, LCB). The data are from Flege (1987a) and Flege and Eefting (1987a).
Figure 8 (top) shows the mean VOT values measured by Flege (1987a) in L1 stops spoken by four groups of subjects who spoke both French and English. None of the bilingual subjects were expected to have established separate phonetic categories for /p,t,k/ in their L2 (French or English) since all of them began learning the L2 in adulthood. This lead to the expectation that they would merge the phonetic properties of similar L1 and L2 stops. For example, the /t/ of L1 would influence the /t/ of L2 and, perhaps to a lesser extent, the /t/ of L2 would affect the L1 /t/. To estimate the effect of L2 learning on L1 production, the bilinguals' production of L1 stops was compared to that of monolinguals.

The results in Figure 8 (top) confirm the existence of an L2 effect on L1 production. Highly experienced French speakers of English (B1, left side) produced French /t/ with longer and therefore more English-like VOT values than French monolinguals. Conversely, highly experienced English speakers of French (Group B3) produced English /t/ with shorter and therefore more French-like VOT values than English monolinguals. These results provided strong support for the hypothesis that equivalence classification leads to the merger of the phonetic properties of pairs of similar L1 and L2 sounds. It would be difficult to explain either the influence of L2 stops on L1 stops or the seeming upper limit on how closely phonetic norms for similar L2 stops can be approximately by even very experienced Late L2 learners if short-lag realizations of /t/ in French and long-lag realizations of /t/ in English were not judged to be equivalent.

The hypothesis offered here is that humans' speech learning ability changes with age because phonetic systems remains sufficiently flexible to permit the establishment of additional phonetic categories for sounds similar but not identical to those already in the phonetic repertoire only up to about the age of 5 to 7 years. When children learning an L2 are exposed to a class of auditorily similar phones unlike any in the L1 (what we have been calling a “similar sound”), they create a new category for the unfamiliar L2 sound. Prima facie support for this hypothesis comes from the observation that the number of phonetic categories established by children in L1 acquisition varies according to the size of the inventory of the L1 being acquired. For example, Spanish-learning children establish far fewer vowel categories than English-learning children. Unlike young children, adults who are exposed to a class of auditorily similar phones unlike any in the L1 tend to reuse existing categories. I call this the “new wine in old bottles” phenomenon.
The data shown in Figure 8 (bottom) are consistent with the dichotomy just drawn between Early and Late L2 learners. This figure shows the mean VOT values in Spanish stops spoken by the Puerto Rican subjects who began learning English L2 as young children (Flege and Eefting 1987a). By hypothesis, they had established additional phonetic categories for English /p,t,k/. A merger of the phonetic properties of corresponding Spanish and English stops, which would have lengthened the VOT in their Spanish /t/s, was therefore not expected. The establishment of long-lag phonetic categories for English /p,t,k/ was expected, instead, to cause shortening of VOT values in their Spanish L1 stops.

The shortening was expected to result from a “polarization” between the early L2 learners’ long-lag categories for English /p,t,k/ and their preexisting short-lag categories for Spanish /p,t,k/. Keating (1984) proposed that the short-lag realizations of Spanish /p,t,k/ have slightly longer VOT values than the short-lag realizations of English /b,d,g/ as the result of a universal phonetic factor called polarization. Polarization serves to enhance the contrast between categories in a phonetic system. In Spanish, the contrast between /p,t,k/ and /b,d,g/ (which are realized with lead VOT values) may be enhanced by increasing slightly the VOT values of the short-lag /p,t,k/s. In English, on the other hand, the contrast between short-lag realizations of English /b,d,g/ and long-lag realizations of /p,t,k/ may be enhanced by a slight shortening of the VOT values for /b,d,g/.

The data in Figure 8 (bottom) suggests indirectly that the Puerto Rican subjects had two phonetic categories to use in implementing the voiceless phonemes /p,t,k/: a long-lag category for voiceless stops in English and a short-lag category for voiceless stops in Spanish. It appears that the two phonetic categories interacted with one another via the mechanism of polarization. The subjects in groups BC, LCB, ECB all produced Spanish /p,t,k/ with shorter VOT values than did age-matched Spanish monolinguals. This is just what one would expect if the long-lag categories established when they began learning English at about the age of 5-6 years put a downward pressure on VOT in their realizations of short-lag (Spanish) phonetic categories.

Based on the studies presented thus far, one might conclude that the tendency for sounds in L1 and L2 to assimilate to one another affects only similar sounds (e.g., Spanish [t]s and English [tʰ]s), and then only for Late L2 learners. There is some reason to question whether demarcation between Early and Late L2 learners should be set at the age of 5-7 years,
however. Flege and Eefting (1987b) showed L2 effects on L1 production 
for some Dutch subjects who had begun learning English at the age of 12 
years. Additional work is needed to better define the dividing line between 
early and late learning.

4. Outline of a Model of L2 Phonological Learning

The experiments reported in this chapter are consistent with the following 
hypotheses which, in effect, outline our current program of research and 
can serve as the basis for a model of the acquisition of L2 phonology. (For 
a more detailed discussion, see Flege 1990a,b).

1. Every human being is born with "phonetic learning ability", i.e., 
the ability to learn to identify the phonetic categories of an input lan­
guage(s) and to produce speech with acoustic properties closely conforming 
to the phonetic characteristics of those categories.

2. L1 phonology "develops" in the following ways:
   (a) Phonetic categories are established before phonemic 
categories. The number of phonemic categories will be smaller 
than the phonetic categories use to implement them when 
phonemes are produced with clearly identifiable allophones 
that are not phonetically conditioned; The number of phonemic 
categories is determined only after a sufficiently large lexicon 
has been established.
   (b) The perceptual representations for phonetic categories are 
elaborated until they conform closely to those of mature speak­
ers in the surrounding community;
   (c) The basic motor plans specified in each phonetic category, and 
the sensorimotor realization rules used to translate the phone­
tic categories into articulatory gestures, are aligned with per­
ceptual representations so that phonemes are produced in a 
language-appropriate manner;
   (d) To varying degrees, all of the above are dependent on the 
quantity and quality of the input received by the learner.

3. Phonetic learning ability remains intact through the lifespan.
Speech learning in L2 differs from L1 speech learning, however, because:
   (a) The phonetic system is gradually optimized for the encoding 
and decoding of the sounds in L1 (as outlined in 2);
(b) The phonetic system becomes resistant to the addition of new phonetic categories, which is partly a result of the optimization process (see Flege 1990a).

4. Whether L2 learners identify phones of L2 as “similar” or “new” has important consequences:
   (a) If similar, learners will substitute sounds from their L1 repertoire;
   (b) If new, learners will eventually produce them independently of sounds in the L1 repertoire.

5. An L1 “accent” in the L2 may result from:
   (a) Immature attempts at a new sound, which may lead to “developmental processes” resembling those of children learning L1 (see Flege and Davidian, 1985);
   (b) Equivalence classification of similar sounds, which may lead to transfer errors in production;
   (c) Incorrect lexical representations (e.g., /lak/ for rock);
   (d) Correct central representations but immature realization rules;
   (e) Some combination of (a)-(d).

6. Age of learning will determine how similar but not new sounds in an L2 are treated:
   (a) Similar sounds are identified increasingly as being inside the phonetic repertoire and less often as being outside the repertoire as age of learning (AOL) increases.
   (b) For individuals who begin learning an L2 before the age of about 5-7 years, additional phonetic categories are established for similar L2 sounds. The corresponding L1 and L2 sounds will be implemented using different phonetic categories and phonetic realization rules. The production of similar L1 and L2 sounds will be authentic.
   (c) For individuals who begin learning an L2 after about the age of 5-7 years, additional phonetic categories will not be established. The corresponding L1 and L2 sounds will mutually influence one another because they are implemented using the same phonetic category. Differences in production may result from the application of different realization rules, but the L2 sounds will not be produced authentically.
   (d) Given sufficient L2 input, L2 learners remain able, even as adults, to establish additional phonetic categories for new L2
sounds. Many of those who do so will go on to produce new L2 sounds authentically. Attitudes and motivation, as well as psychosocial factors may play a role in defining phonetic input in these instances, and so may have an impact on how well new L2 sounds are produced.

5. Future Directions

The hypotheses formulated above will need to be further tested, and other areas of L2 phonology explored. It should be noted that at present, for example, the model makes no prediction concerning what kind of L1 phonological rules, if any, will be applied to the production of words in an L2 (see, e.g., Rubach 1984, for a discussion of cyclic versus post-cyclic rules). The model does not now deal with intonational and prosodic characteristics of L2 speech, although it is clear that such phenomena are important (see e.g., Willems 1982; Flege and Bohn 1989a).

As mentioned earlier, a priority in the area of L2 phonological research must be the development of a technique for classifying sounds in an L2 as new or similar. A decade ago Wode (1978) noted that phonetic theory did not provide a metric that could be used to determine when the equivalence classification of phones in L1 and L2 will occur. Although some advances have been made (see Flege 1990a), and new approaches are currently being considered in our laboratory, this is still the case.

Our work has suggested the existence of both phonemic and phonetic levels of organization in both the encoding and decoding of speech (see also Werker and Logan 1985; and Flege 1990a). In defining the "sounds" of an L2, it will be necessary to draw a distinction between phonological and phonetic equivalence. For example, the /l/s in two languages may be equivalent at a phonological level, in the sense that they occupy the same place in a vowel system, but be non-equivalent phonetically (Wode 1978).

One theoretical and methodological problem facing the speech learning model sketched here (and presented in more detail in Flege 1990a) is the problem of individual subject variability. In recent VOT studies, we have noted a small subset of adult L2 learners who have defied the predictions of the model by producing similar English stops with native-like VOT values (see Underbakke et al. (1988), for an analogous observation about
Japanese subjects' perception of the English /r/-/l/ contrast). Also contrary to the model, certain adult L2 learners appear unable to produce new L2 sounds authentically.

An example of the apparent inability of adults to learn a new L2 sound can be found in Altenberg and Vago (1987). These authors reported that two highly experienced Hungarian speakers of English substituted Hungarian /i/ for English /i/. A comparison of the phonemic inventories of Hungarian and English suggests that English /i/ is a new vowel. What, then, prevented the Hungarians from producing English /i/ authentically? One possibility is that English /i/ is not actually a new sound, as suggested by a cross-language comparison of phonetic symbols (see Section 3.1). Perhaps Hungarians tend to regard the contrast between English /i/ and /i/ as being analogous to that between Hungarian /i:/ and /i/, leading them to equate Hungarian /i/ with English /i/. If so, it means that phonological as well as phonetic structure is important to the process of L2 phonetic learning.

If this last observation is correct, it would mean that English /i/ is a similar vowel sound for Hungarians. This should have two measureable effects: incomplete approximation to English phonetic norms for /i/ in a carefully controlled instrumental study, and a gradual movement of Hungarian /i/ towards English /i/ in the acoustic phonetic vowel space. Flege (1990b) discussed a number of factors that might lead to individual subject differences, and proposed that inter-subject variability increases beyond the age of 5-7 years due to an increasing likelihood that those factors will affect L2 speech production.

Future research will be needed to account for individual subject differences. One wonders, for example, if important differences exist in the way various L2 learners process L2 sounds auditorily or perceptually. Is the effect of equivalence classification weaker for certain individuals because they are better able to process phones in a phonetic rather than a phonemic mode? Are certain learners able to mimic unfamiliar sounds better than others because they are able to suspend the process of covert phonemic categorization?

As discussed by Flege (1987b), valid tests of a sensitive period hypothesis are difficult if not impossible to achieve. This is because of the enormous difficulty in controlling for factors other than age of learning (AOL). It is difficult to quantify, for example, the quantity and quality of phonetic input received by the L2 learner. Important conclusions have been drawn in this chapter about the nature of phonetic learning based on the
inference that certain subjects received accented input as they learned L2. In reality, it is difficult to establish accurately what subjects have actually heard as they learned L2. Other important conclusions were based on inferences concerning the age at which L2 learning began (i.e., AOL). But, again, it is difficult to know exactly when L2 learning commenced and, in some instances, whether the nominal L2 is really a second language (see McLaughlin 1978).

These uncertainties pose a methodological challenge to L2 researchers. It would be desirable in future studies to obtain more accurate information about individual subjects' actual history of phonetic input and language use. The popular strategy of comparing extreme groups may need to be supplemented by laboratory training studies in which phonetic input is controlled carefully. Such an approach may provide a viable alternative to cross-sectional, observational studies as a means for studying changes in the human speech learning ability across the lifespan.

In conclusion, the data presented here suggest that foreign accent does not arise from an irreversible decrease of speech learning ability brought about by the passing of a critical period. It may instead result from the development of the L1 phonetic system, which makes it increasingly unlikely that similar sounds in an L2 will evade being equated with sounds in L1. If the model is supported, it may motivate speech-language pathologists and language teachers to develop and implement more effective methods with which to teach pronunciation. Two studies have already provided interesting possibilities for future development: the use of visual feedback concerning the tongue's position (Flege 1989a), and feedback concerning phonetic category identity in perceptual training experiments (Flege 1989b; Flege and Wang 1990).

Notes

1. This chapter was supported by NIDCD grant NS20963. The author thanks C. Ferguson and T. Huebner for editorial contributions; O.-S. Bohn, G. L. Menn, and Yeni-Komshian for comments on an earlier draft of the chapter; and those who have contributed to our ongoing program of research on speech learning: O.-S. Bohn, W. Eefting, R. Hammond, J. Hillenbrand, R. Major, and C. Wang.

2. Somewhat surprisingly, the bilingual children (Group BC) did not differ significantly from a group of monolingual English children even though they, like the adults in ECB and LCB, produced English /p,t,k/ with shorter VOT values than age-matched English
monolinguals. This last finding suggests that the perception of English stops by the native Spanish children was more nearly native-like than their production of English stops. It suggests, further, that although the Spanish children had English-like perceptual representations for /t/, they had not yet learned to output stops in the manner specified in those representations.

3. Flege (1990b) suggested that the stage in speech development at which it becomes more likely for an individual to assimilate a similar L2 sound to an existing L1 category rather than to establish a new category for it occurs at about the age of 5-7 years. This suggestion was based on the estimated age of L2 learning at which a foreign accent may first be evident.

References


PERCEPTION AND PRODUCTION


Perception.” *Perception & Psychophysics* 37.35-44.
*Child Phonology, Volume 2, Perception* ed. by G. Yeni-Kenshian, J. Kavanagh and