PHONOLOGICAL DEVELOPMENT
Models, Research, Implications

edited by
Charles A. Ferguson
Lise Menn
Carol Stoel-Gammon
Chapter • 21

Speech Learning in a Second Language

James Emil Flege

Adults and older children learning the sound system of a second language (L2) differ from young children acquiring the sound system of their native language (L1) in two important respects. The L2 learners are better able to control their speech apparatus than the young children acquiring an L1, and they already possess a phonetic system for producing speech. As a result, far more errors in production are likely to arise from the inappropriate use of previously acquired structures in L2 learning than in L1 acquisition. It is well known that adult learners are rarely, if ever, completely successful at mastering the sound system of an L2. A widely held view is that when adults encounter an L2 word, they attempt to “decompose” it into the phonemic units of the L1, and then produce the L2 word as if it consisted of phonetic elements (allophones, phonemes) from the L1 (Polivanov 1931).

This chapter provides a brief overview of research that has de-

---

1We use the term L2 learning rather than acquisition because of the view that phonetic systems, even those of adults, undergo constant change in the face of new phonetic input. Thus, speech is never fully acquired. The term speech learning as used here refers to all aspects of learning that affect the production and perception of the sounds making up words. It is used in preference to phonological learning because much of our research to date has focused on phonetic-level processes.

The preparation of this chapter was supported by NIH grant DC00257. The author thanks colleagues (O.-S. Bohn, W. Eefting, M. Munro), graduate students (K. Fletcher, S. Jang, C. Wang), and research assistants (L. Cueva, C. Mena, L. Skelton) who have participated in work reported here. Thanks are also extended to O.-S. Bohn, C. Ferguson, and A. Walley for editorial comments.
scribed L2 learners' production and perception of L2 sounds. The view that foreign accent is based on the substitution of the nearest LI sound for the sounds making up L2 words appears to be inadequate. Such patterns do occur frequently, especially in the earliest stages of L2 learning. Examples of what might be called cross-language "phonetic interference" are presented. However, most adults eventually modify previously established patterns of speech production and perception when attempting to deal with the sound system of an L2. Thus, the differences one might observe between native and non-native speakers often involve more than just the maintenance of old articulatory habits. A recurrent theme of the chapter is that many aspects of L2 production sound may be produced authentically if it is identical to a sound in the LI, or if it is so similar to an L1 sound that the differences between it and the nearest L1 sound will go unnoticed if the L1 sound is substituted for the L2 sound. We have hypothesized that L2 sounds that differ substantially from any sound in the L1 may also be produced authentically, at least once L2 learners have received sufficient phonetic input. The most serious learning difficulty may occur for L2 sounds that are different enough from any LI sound that L1-for-L2 substitutions could be noticed readily, but not so different as to trigger the formation of new phonetic categories for the L2 sounds.

PRODUCTION VERSUS PERCEPTION

According to the contrastive analysis (henceforth, CA) approach of the 1950s and 1960s, cross-language differences were the primary source of speech learning difficulty. L2 sounds were expected to be produced authentically (i.e., accurately) by L2 learners if they closely matched LI sounds, but poorly if they did not. This implied that the authenticity of L2 production would depend on the magnitude of the phonological differences between L2 sounds and the closest sound(s) in the LI. However, a "U" shaped rather than a linear function may describe better the effect of varying differences between L1 and L2 sounds. An L2 sound may be produced authentically if it is identical to a sound in the L1, or if it is so similar to an L1 sound that the differences between it and the nearest L1 sound will go unnoticed if the L1 sound is substituted for the L2 sound. We have hypothesized that L2 sounds that differ substantially from any sound in the L1 may also be produced authentically, at least once L2 learners have received sufficient phonetic input. The most serious learning difficulty may occur for L2 sounds that are different enough from any LI sound that L1-for-L2 substitutions could be noticed readily, but not so different as to trigger the formation of new phonetic categories for the L2 sounds.

CA came into disfavor largely because of its failure to predict which particular L2 sounds would, or would not, be difficult. In retrospect, more serious problems existed. First, despite evidence for important differences in phonological and phonetic levels of encoding and decoding (see e.g., Flege 1991 b, c, 1992a), CA did not usually differentiate levels of processing. An attempt was simply made to match the phonemes of the LI and L2. Second, CA never made it clear whether the difficulties encountered by L2 learners had a primarily motoric or perceptual basis.

It has been suggested that L2 production errors arise because the ability to learn new forms of pronunciation diminishes with age (Sapon 1952). Difficulties in L2 pronunciation could arise from an inability to modify previously established patterns of segmental production or to develop new ones. For example, neither English /r/ nor /l/ is realized with phones that occur on the phonetic surface in Japanese. Not surprisingly, Japanese learners often have difficulty in producing English /r/ and /l/ correctly. They are judged by native speakers of English to make more production errors for word-final singletons than for word-initial singletons, and more errors for word-initial singletons than for word-initial clusters. Such errors probably cannot be attributed solely to inaccurate speech perception. This is because Japanese learners have more difficulty in identifying English /r/ and /l/ tokens that occur in word-initial clusters than in word-final singletons, and more difficulty for word-initial than for word-final singletons (see Flege 1988a, p. 328 ff).

L2 production difficulties also have been noted for L2 sounds that do have a counterpart in the LI inventory, but that occur in an unfamiliar phonetic context or position. For example, Spanish learners appear to be less successful in producing English /s/ in word-final than word-initial position, just the reverse of the pattern seen for children acquiring English as an LI (Titiritz 1981, cited by Flege 1988a). This pattern seems to arise because Spanish has few word-final consonants. The word-final fricatives that Spanish does have may be weakened or omitted in production, something that might be regarded as a "natural" process. Similarly, Chinese speakers have great difficulty producing an effective contrast between English /p,t,k/ and /b,d,g/ in word-final position, even when their LI has a voicing contrast much like that of English in the word-initial position. Flege, McCutcheon, and Smith (1997) found that Chinese learners of English did not sustain closure voicing in word-final /b/ to the same extent as native speakers of English. This apparently occurred because the Chinese subjects failed to expand the oral cavity actively during labial closure. And, unlike native speakers, the Chinese learners did not produce /p/ with a more forceful labial closure than /b/ (Flege 1988b).

---

2The term sound is used here to refer to a class of phones that can be used to contrast meaning. It is used as a terminological convenience to avoid the constant need for drawing a distinction between the phonetic and phonemic levels. For an exposition of this theoretically important distinction, see Flege (1992a).
These differences between native speakers of English and the Chinese subjects probably did not arise from a lack of awareness that /p/ and /b/ contrast phonemically. Chinese subjects in the Flege, McCutcheon, and Smith (1987) study resembled English speakers in producing /p/ with greater peak oral air pressure than /b/. This suggested a difference in laryngeal articulation. The Chinese subjects also made vowels significantly longer before /b/ than /p/, although the magnitude of their voicing effect was considerably smaller than that of the native English subjects’ (see also Flege, Munro, and Skelton 1992).

There are many other instances, however, in which a difference in production between native and non-native speakers can probably be traced to an underlying perceptual difference. To take an example, Hindi has dental and retroflex stops, but it does not possess stops like English /t/, which are characteristically produced with an alveolar place of constriction (Dixit and Flege 1991). Hindi speakers are reported to use Hindi retroflex rather than dental stops in producing English /t/ (Ohala 1978) even though retroflex stops are considered to be “marked” because of their comparative rarity. This may mean that English /t/ is perceptually more similar to the Hindi retroflex stops and that, in some instances, perceptual similarity is a more important determinant of L1-for-L2 substitution patterns than is articulatory similarity (or difficulty).

A number of studies have demonstrated the existence of differences in phonetic perception between native and non-native speakers (e.g., Miyawaki et al. 1975). Flege and Hillenbrand (1985) showed that native speakers of Swedish and Finnish differentially labelled the members of a synthetic /pe/-/pa/ (i.e., /sl/-/z/) continuum. In this sense, they resembled native speakers of English, but they did not use fricative duration in an English-like fashion. This particular cross-language perceptual difference might be attributed to the absence of final /sl/-/z/ contrasts in Swedish and Finnish, but there is also ample evidence of differences for L2 sounds that do have a phonological counterpart in the L1. For example, the phoneme boundaries of native English and Spanish subjects differ for a /d/-/l/-/l/ continuum in which VOT (voice onset time) has been varied (Flege and Eefting 1986). This difference reflects how /p/, /l/, and /l/ are produced in English and Spanish. It suggests that phonetic perception is attuned to the acoustic characteristics of L1 stops so as to optimize perception.

Cross-language perception differences may reflect more subtle, multidimensional cues. For example, French learners of English might give greater weight to release burst cues in word-final stops than native speakers of English because French stops, unlike English final stops, are usually produced with an audible release burst (Flege and Hillenbrand 1987). The Chinese subjects mentioned earlier may have failed to produce English /p/ and /b/ authentically because their perceptual representations for these sounds differed from native English speakers’. For example, they may have produced only a small difference in the duration of vowels preceding /b/ and /p/ because their perceptual representations did not encode a large-magnitude difference. Flege (1989b) showed that Chinese speakers of English can accurately identify word-final tokens of /t/ and /d/ when they contain release bursts, but do so more poorly than native speakers when final release bursts and closure voicing cues are removed, even after feedback training (Flege and Wang 1990). This suggests that Chinese learners attempt to apply the representations used to identify Chinese word-initial stops, which are released, to English word-final stops, which are often unreleased. If so, then the relevant level of analysis for describing L2 perception difficulties may be the phonetic rather than the phonemic level (Brière 1966).

Sheldon and Strange (1982) showed that Japanese learners’ production of English /t/ and /d/ may be more native-like than their perception of these unfamiliar sounds. Such a finding has not occurred frequently in either the L2 or the L1 literature. It may have arisen from the application of conscious articulation strategies in the context of a formal experiment. It has been hypothesized (e.g., Flege 1981) that accurate phonetic perception is a necessary but not sufficient condition for accurate L2 segmental production. On this view, the Spanish learners mentioned earlier may have omitted word-final fricatives more often than native English speakers because their perceptual representations of word-final fricatives differed from those of native English speakers in terms of how the multiple acoustic dimensions that specify /s/ are integrated. Or, they may have differed from native speakers at a more abstract, phonological level in terms of the hierarchical arrangement of distinctive features (see Weinberger 1990).

The possibility exists, of course, that the Spanish adults’ perceptual representations of final /s/ closely resembled those of native English speakers, and that their production difficulty was purely motoric. A speech production difficulty for final /s/ could be localized at a segmental level, or derive from an overall difference in syllable structure between Spanish and English (see Delattre 1951). These are empirical questions. Unfortunately, the L2 literature has seldom resolved such questions because few studies examining both the production and perception of L2 sounds have been carried out. Even with the availability of parallel production-perception data sets, it will likely continue to prove difficult to resolve questions concerning the contribution of perceptual and motoric factors to errors in L2 production because of difficulty in finding a common metric for gauging the distance between L1 and L2 sounds in the two domains.
INTERLINGUAL IDENTIFICATION

It is generally assumed that L2 learners identify certain L2 and L1 sounds with one another (Weinreich 1953; Wode 1978, 1981). This process, which is based on judgments of overall phonetic, and perhaps phonological, similarity appears to occur even when the acoustic difference between the L1 and L2 sounds is detectable auditorily. When an L1 sound is identified with a sound in L2, it may be used in place of the L2 sound (Valdman 1976). The use of multiple L1 substitutes for an unfamiliar L2 sound, especially in early stages of learning (Hammarberg 1990), suggests that in some instances an unfamiliar L2 sound may be judged to be perceptually equidistant from two or more L1 sounds.

It has often been claimed that the phonology of L1 causes L2 learners to “filter out” acoustic differences that are not phonemically relevant in the L1 (e.g., Trubetzkoy 1939; Wenk 1985). Flege, Munro, and Fox (1992) examined the perceived dissimilarity of pairs of vowels drawn from Spanish and/or English. Multidimensional scaling (MDS) analyses showed that native Spanish-speaking listeners used fewer dimensions in judging between vowel dissimilarity than native English-speaking listeners. This is analogous to the finding that Spanish talkers use a narrower range of tongue positions to produce Spanish /i, a, u/ than English speakers use in producing English /i, a, u/ (Flege 1989a). Listeners whose L1 has a small vowel inventory may use fewer dimensions in identifying English vowels than native speakers of English because there are fewer phonetic contrasts to maintain in the L1.

Flege, Munro, and Fox (1992) found that native Spanish listeners judged certain pairs of phonetically (and phonemically) distinct English vowels to be less dissimilar than native English listeners, apparently because they did not generate two distinct phonetic codes for the vowel pairs. However, pairs that were likely to have been labelled differentially by both English and Spanish listeners often received much the same dissimilarity ratings. For example, both listener groups judged /i/-/a/ pairs to be more dissimilar than /i/-/e/ pairs. Such agreement implies the existence of a universal, auditory-based metric for similarity judgments. Universal “phonetic learning strategies” may also exist. Bohn and Flege (1990a, b; Flege and Bohn 1992) found that inexperienced German and Spanish learners of English relied on duration when asked to identify the members of synthetic vowel continua whose endpoints did not contrast in their L1. Reliance on duration could not

———

This finding agrees with the observation that large-inventory languages will tend to exploit a wider range of articulations than small-inventory languages to ensure a sufficient degree of perceptual contrast (Lindblom 1988).

be attributed to previously learned (i.e., L1) perceptual patterns, at least not for the Spanish subjects. It was, therefore, concluded that L2 learners may exploit temporal cues to a greater extent than spectral cues to identify unfamiliar vowels in a second or foreign language, at least initially.

It is nevertheless true that L1 background may influence importantly the perceived similarity of sounds drawn from two languages (Butcher 1976). Results obtained by Flege and Wang (1990) suggested that learned patterns of attentional allocation may be carried over from the L1 into an L2. Gottfried and Beddor (1988) found that American subjects used duration cues to a greater extent than French subjects in identifying the members of a synthetic French /œ/-/œ/ continuum. This was attributed to a greater overall use of duration cues in English than French. Munro (1990) found that native speakers of a language with a phonemic length distinction (viz. Arabic) used duration to a greater extent to classify vowels from an unknown foreign language than native speakers of English.

Caldirop asserted (1968, p. 164) that the only basis for interlingual identification is “substantial . . . rather than formal.” It may be difficult to draw a clear distinction between the two, however. For example, the finding by Munro (1990) implies that if native speakers of English and Arabic were asked to judge a pair of vowels differing greatly in duration, the Arabic subjects would rate it as more dissimilar than the English subjects. That is, different judgments of cross-linguistic similarity may be rendered by speakers of different languages because of differences in the prominence given to certain acoustic dimensions, or to the way distinctive features are arranged hierarchically in particular languages (Clements 1985).

There is evidence that speech is processed at auditory, phonetic, and phonemic levels. Listeners’ conscious awareness of speech is usually a reflection of phonemic-level processing, but listeners can gain access to the phonetic level under certain circumstances. Flege (1991d) hypothesized that the sound systems of two languages may interface at a phonetic as well as at a phonemic level. This was supported by a study in which Spanish and English subjects estimated degree of rhyme in English words (Flege 1992b). Spanish and English subjects rated pairs of English words with /e/ and /e/ (e.g., mate-melt) to rhyme equally, but gave different ratings for other word pairs. This finding suggests that at least one kind of judgment of intervowel distances, viz. rhyming judgments, may be made at a phonetic category level. Had the rhyming judgments been made at a phonemic level, the Spanish subjects should have perceived a greater degree of rhyme than the English subjects, for the Spanish /e/ phoneme is realized with [e] and [ɛ] allophones (see also Flege 1991a).
This finding is not conclusive, however, because of the possibility that the Spanish subjects identified the English /æ/ in terms of the /eɪ/ sequence of Spanish. It seems to agree, however, with results obtained in another recent study using a different technique. Flege, Munro, and Fox (1992) had Spanish and English listeners rate consonant-vowel (CV) pairs containing English /æ/ and Spanish /eɪ/ realizations for degree of dissimilarity. If the ratings had been based on phonemic-level representations, the /eɪ/-/æ/) pairs should have been judged to be less dissimilar by the Spanish than the English listeners. This is because pairs of vowels that are given the same label are generally rated as more similar than vowel pairs labelled differently. The English and Spanish listeners did not differ, however, suggesting that the comparison was not based on phonemic codes. That is, the Spanish listeners appear to have based their judgment on the phonetic quality of the /eɪ/ tokens, even though the allophone /æ/ of Spanish /eɪ/ is not permitted to occur in open syllables. Admittedly, however, the role of phonetic-level processing in interlingual judgments will have to be examined further.

NEW VERSUS SIMILAR SOUNDS

Some sounds in an L2 may be identical to those found in the L1. Most L2 sounds differ from L1 sounds, however, if only in minor details of timing, amplitude, or placement. An important issue is whether interlingual identification persists for all nonidentical L2 sounds after L2 learners have become familiar with the sound system of the L2. We have hypothesized that certain “new” L2 sounds that differ substantially from any vowel in the L1 will cease being identified with a sound(s) in the L1 inventory. “Similar” sounds that more closely resemble a sound in the L1 inventory, on the other hand, will continue to be identified with an L1 sound (i.e., be “equated”).

Equivalence Classification

If L2 learners persist in identifying an L2 sound with an auditorily distinct sound in the L1 inventory, it is said to have been “equated” with the L1 sound. Equivalence classification is of undoubted importance during L1 development, for it permits the child, and even infants (Kuhl 1980; Hillenbrand 1983), to group disparate phones into functional categories and to perceive constancy in the face of acoustic phonetic variation due to factors such as talker gender, stress, and speaking rate. Flege et al. (1992) provided evidence supporting the effect of equivalence classification in L2 learning. Spanish subjects who were relatively experienced and inexperienced in English gave much the same dissimilarity ratings for vowel pairs consisting of a Spanish /i/ token and an English /i/ token. This finding suggested that the English /i/ did not emerge as a “new” vowel for the experienced Spanish speakers of English, which is hardly surprising because the Spanish /i/ is only slightly lower in an acoustic-phonetic and articulatory space than its English counterpart (Flege 1989a). It was apparent, however, that the Spanish subjects could distinguish the /i/ of Spanish and English auditorily. Both the native Spanish and English subjects rated pairs consisting of a Spanish /a/ token and an English /a/ token as being more dissimilar than pairs consisting of a Spanish /e/ and a Spanish /i/ token.

Defining the New Versus Similar Distinction

Students of bilingualism have long drawn a distinction between what we now call new and similar sounds (Delattre 1964, 1969; Bridère 1966). However, at present no standard method exists for determining if an L2 sound will be treated as new or similar. The distinction is based on differences in the perceived phonetic distance between sounds in the L2 and those in the L1 (see Best 1990). Although both new and similar L2 sounds differ acoustically from sounds in the L1, there is thought to be a qualitative as well as a quantitative difference in the degree of phonetic dissimilarity between L1 sounds and sounds in the L2 that are treated as new and similar. Wode (1978), for example, supposed that there is a threshold beyond which an L2 sound will be recognized as distinct from any sound in the L1.

A reliable metric for classifying L2 sounds as new or similar must be developed if predictions concerning these two supposedly different L2 sound types can be tested. Flege (1992a) discussed methods that might be used to operationalize the distinction. An L2 sound may be defined as similar if it is represented by the same IPA symbol as a sound in the L1, provided it can be shown to differ auditorily from the corresponding L1 sound. New sounds, on the other hand, might be defined as L2 sounds that are represented by an IPA symbol that is not used to represent a sound in the L1 (and, of course, which differs au-

4In Spanish, words like renato ([reina] “queen”) contrast with words like renar ([renar), “reindeer”). English learners of Spanish have difficulty distinguishing Spanish /eɪ/ and /æ/ (Ferguson 1990), apparently because they are accustomed to the absence of an offglide for English /æ/ at fast speaking rates. The /æ/ in many English words borrowed by Japanese, a language whose vowel system resembles that of Spanish, has been rendered as a two-vowel /eɪ/ sequence (Lovins 1975). It is therefore possible that pairs like mate–mato in the Flege (1992b) study rhymed little for the Spanish subjects because they interpreted the English word mate to have the two-vowel sequence /eɪ/ and mato to have /æ/. Whether or not the extent of formant movement in English /æ/ is sufficient to cue the existence of a two-vowel sequence for Spanish speakers is an empirical question that needs to be resolved.

5It is uncertain if perceived articulatory similarity enters into cross-language judgments of phonetic similarity.
editorily from the nearest LI sound). Problems with this approach do exist, however. This is especially true for the “phonetic symbol” test, for transcriptions may vary according to individual practice and ability (see Flege 1992a).

Alternatively, an L2 sound could be classified as similar if (1) it were shown to be phonetically distinct from LI sounds using one of several “phonetic distinctness” tests, and (2) it could be shown to differ auditorily from the nearest LI sound(s). Native speakers of a language having a phonemic writing system, such as Spanish or Finnish, could be asked to label L2 vowels with the letters used to write vowels in their LI, or to respond “none” if they heard a vowel not found in the LI (see Flege 1991a). L2 vowels that received the “none” label frequently would be considered phonetically distinct from vowels in the LI inventory. Other phonetic distinctness criteria might also be used, such as rhyming judgments. So a new L2 vowel might be defined as one that rhymes less with any vowel in the LI than does any pair of vowels drawn from two adjacent LI vowel categories. Dissimilarity ratings might also be used as a means for assessing the status of L2 sounds. A new sound might be defined as an L2 sound that is judged to be more dissimilar when paired with realizations of the closest LI category than pairs of sounds drawn from that LI category and its closest neighbor.

Vowel Production

The CA approach leads to the view that the rate and eventual extent of learning for various L2 sounds will not differ. However, as predicted by our model of L2 speech learning, evidence is accumulating that certain L2 sounds are learned more readily than others. Pimsleur (1963) found that discrimination training led to an improved pronunciation of new French vowels but not of the “similar” French vowel /æ/. Results reported by Henning (1966) suggested that native English subjects who did not speak French were better able to imitate French vowels that partially resemble English vowels than new French vowels. This suggests that, at least initially, the partial resemblance of an L2 sound to a sound in the LI may prove helpful. However, Mueller and Niedzielski (1963) found just the opposite pattern in a study examining the spontaneous productions of students enrolled in a French class. They were judged by a native French-speaking listener to have produced new French vowels (e.g., /y/) far better than similar vowels (e.g., /æ/). These results for French were verified in an instrumental phonetic study (Flege 1987b)

examining the production of French vowels by native speakers of French, and by native English speakers of French who had resided in Paris for 12 years. Acoustic analysis showed that the native English subjects produced French /y/ authentically. This vowel has no phonological counterpart in English. The native English subjects differed from native French speakers, on the other hand, in producing the similar French vowel /u/, which is not fronted (or made into a diphthong) like its English counterpart.

Several recent studies have provided converging evidence that adult learners are able to master /æ/ if their LI does not have such a vowel. Perceptual results obtained by Major (1987a, see also Flege 1992a) suggested that as Brazilians’ global foreign accent in English improved, the identifiability of their English /æ/ increased. This did not hold true for the production of English /æ/, a similar vowel with a counterpart in Portuguese. In an acoustic analysis, Bohn and Flege (1990c, 1992a) found that inexperienced but not experienced German L2 learners differed from native speakers of English in producing the new vowel /æ/.

Flege (1992a) provided perceptual evidence that Dutch learners with a poor pronunciation of English, but not those with a good overall pronunciation of the L2, differed from native English speakers in producing the new English vowel /æ/. L2 learners in both the German and Dutch studies showed small but measurable differences from native speakers for similar English vowels known to differ acoustically from corresponding vowels in the LI. This was predicted from the hypothesis (e.g., Flege 1987b) that L2 learners are unable to establish additional phonetic categories for similar L2 vowels because they are equated with LI vowels.

L2 Consonant Production

Best, McRoberts, and Sithole (1988) provided perceptual evidence suggesting that Bantu clicks may be sufficiently distinct from any English consonant that they evade equivalence classification by native English speakers during the process of L2 learning. However, no L2 production study to my knowledge has examined the learning of clicks or other so-called “exotic” sounds by native English speakers.7

Quite a few studies have examined the production of English /l/

7In the literature written in English, the term exotic usually refers to non-English sounds that occur rarely in human languages. Such sounds are likely to be treated as new by native English speakers but, because they are rare, may be articulatorily complex or difficult. To test adequately the hypothesis that adults can master new consonants in an L2, it will therefore be necessary to find L2 consonants that are not known to be difficult articulatorily as shown by, for example, relatively late mastery during LI acquisition.
Other studies suggested that the production of English /t/ and /l/ may approach native-like levels. Dissoway-Huff (1981) examined minimally-paired English words spoken by three beginning-level Japanese students. Correct identification rates by native English listeners were 80% (word-initial), 67% (intervocalic), and 67% (word-final). Mochizuki (1981) reported an average 92% correct identification rate for a single Japanese subject reported to have a poor command of English. Sheldon and Strange (1982) examined words spoken by Japanese subjects who had lived in the U.S. for an average of 1.8 years. The correct identification rate for /t/ and /l/ was just 65% for one talker, but it averaged 99% for the other five talkers examined. Although this last finding suggests that English /t/ and /l/ can be mastered, a definite conclusion would be premature. The words examined by Sheldon and Strange (1982) were read from a list, then assessed in a forced-choice identification experiment. There is no guarantee that Japanese learners' /t/ and /l/ attempts will be judged to be fully acceptable (i.e., undistorted) in conversations, or even that their production of these English consonants will be as readily identifiable. Perhaps more importantly, the data obtained by Sheldon and Strange (1982) for /t/-/l/ production may not generalize to the majority of Japanese learners of English, for their subjects were apparently selected on the basis of pronunciation proficiency. Although English /t/-/l/ learning has been studied often, more work is needed.

Much L2 acoustic research has focused on the voice onset time (VOT) dimension. English /p/, /t/, and /k/ have been examined in several studies. These phonologically voiceless stops are produced with long-lag VOT values. Flege and Hillenbrand (1984; Flege 1987b) hypothesized that even highly experienced French adults would equate realizations of English /p/, /t/, /k/ with their counterparts in French, even though the acoustic difference between short-lag and long-lag realization of /p/, /t/, /k/ appear to be detectable auditorily (Flege and Hammond 1982, Flege 1984, 1991b). This led to the prediction that French adults would be unable to establish phonetic categories for English /p/, /t/, /k/ as the result of equivalence classification, and would, thus, fail to produce English /p/, /t/, /k/ authentically.8

As predicted, the French adults examined by Flege and Hillenbrand (1984) did not produce English /p/, /t/, /k/ authentically. Many other L2 production studies with subjects whose L1 has short-lag /p/, /t/, /k/ have shown the same thing. Adult L2 learners tend to produce English /p/, /t/, /k/ with short-lag VOT values, or with compromise values that are intermediate to the VOT norm for /p/, /t/, /k/ in the L1 and L2 (Flege and Port 1981; Port and Mitleb 1980, 1983; Nathan 1987; Flege and Eefting 1987a; Major 1987b; Lowie 1988). Only a few individual exceptions to this general rule have been noted. The fact that experienced L2 learners produce /p/, /t/, /k/ with significantly longer VOT values in English than in their L1 demonstrates that they have detected at least some of the phonetic differences between /p/, /t/, /k/ in L1 and L2. If late learners did not equate the two kinds of voiceless stops, it would be hard to understand why they so seldom manage to differentiate fully the L1 and L2 versions of /p/, /t/, /k/.

Another kind of evidence supporting the hypothesis that long-lag voiceless stops in L2 will be equated with the short-lag voiceless stops in L1 is the existence of L2 effects on L1 stop production. If foreign accent were simply a matter of "interference" one would expect to see effects of the L1 phonetic system on L2 production, but probably not the reverse. However, Flege (1987b) found that Americans who were highly experienced speakers of French produced English /t/ with shorter, and thus slightly more French-like, VOT values than English monolinguals. (The reverse pattern held true for highly experienced French speakers of English; they produced English-like stops in French.) Flege and Hillenbrand (1984) hypothesized that when similar L1 and L2 stops are equated, a learner's L1 voiceless stop category will be restructured so as to accommodate the range of L1 and L2 stops that have been identified as realizations of it.

EARLY VERSUS LATE LEARNERS

In support of this, Major (1990) provided preliminary evidence that the more closely adult learners approximate the L2 VOT norm for voiceless stop consonants, the more their L1 stops will come to resemble L2

---

8The claim that French and English /p/, /t/, /k/ realizations will be equated might at first seem counterintuitive. Many speech perception experiments with synthetic VOT continuoa have shown that subjects whose L1s have short-lag, and long-lag versions /p/, /t/, /k/ will differ in the location of their voiced voiceless phoneme boundaries. However, recent work with natural stimuli has shown that VOT is not an overriding cue to stop voicing (Forrest and Rockman 1988). Bohn and Flege (1991b) found that, despite the well-known VOT difference between the /p/, /t/, /k/ of Spanish and English, Spanish monolinguals consistently identified long-lag English /t/ tokens as /l/. In a surprisingly large number of instances (52%), Spanish /t/ tokens were identified by English monolinguals as /t/. Clearly, short-lag Spanish /t/ tokens have a property (or properties) that make them sound like phonologically voiceless stops to native speakers of English.
stops, at least in casual speech (see also Flege 1988a). Major suggested that adult learners may be unable to maintain “two language systems with native-like fluency” because of the mutual effects of L1 and L2. Psycholinguists have claimed, for different reasons, that the two language systems of bilinguals cannot operate completely independently of one another because the L1 and L2 systems are always co-activated to some degree (Obler and Albert 1978; Grosjean 1982, 1985, 1989; Grosjean and Soares 1986).

The seeming upper limit on how authentically English /p,t,k/ can be produced may apply to children and adolescents as well as to adult L2 learners. The strength of global foreign accent in sentences is known to be inversely related to the authenticity of VOT in English /p,t,k/ (Flege and Eefting 1987b; Major 1987a). Sometimes even young children may speak their L2 with an accent (e.g., Asher and Garcia 1969). This leads one to expect that children may not produce the VOT in L2 stops correctly. This inference has, in fact, been supported by studies examining adults who learned English as children or adolescents (Williams 1979, 1980; Suomi 1980; Flege and Eefting 1987b; Schmidt 1988). Caramazza et al. (1973) found that native French speakers who began learning English in Canada by the age of seven years produced English /p,t,k/ with “compromise” VOT values that were intermediate between the values typical for the L1 and L2. Flege and Eefting (1987b) found much the same for native Spanish subjects in Puerto Rico who began learning English by the age of five to six years. (Similar compromise values were observed for early learners who were adults and children at the time of testing.)

These findings raise the issue of whether anyone can learn to produce /p,t,k/ authentically in an L2. It appears, however, that English /p,t,k/ can be mastered if L2 learning commences by the age of five to six years. Flege (1991c) examined L1 and L2 stop production by early learners who received native-speaker input when they first began learning English at school in Texas at the age of five to six years. These early learners closely resembled Spanish monolinguals in producing Spanish /t/, and English monolinguals in producing English /t/. That is, they produced a full phonetic contrast between /t/ in Spanish and English. Late learners, on the other hand, produced English /t/ with the expected compromise VOT values. It was hypothesized that individuals who begin learning the L2 by about the age of five to six years can accurately produce similar L2 sounds because they are able to establish separate phonetic categories for the corresponding L1 and L2 sounds (see Flege 1988a).

It may be the case that the early learners examined by Flege and Eefting (1987b) and by Caramazza et al. (1973) received accented L2 input. If so, they may have formed phonetic categories that specified the compromise VOT values they had heard as young children. This inference was confirmed by a study of the Puerto Rican subjects examined earlier by Flege and Eefting (1987b). In a study by Flege and Eefting (1988), these subjects imitated the members of a synthetic continuum whose members spanned all three modal VOT categories (i.e., lead, short-lag, long-lag). Spanish monolinguals, English monolinguals, and late L2 learners tended to produce stops with VOT values falling into just two of the three modal VOT categories. Early learners, on the other hand, used all three modal categories. This suggested that they had two pitch categories that might be used to implement /t/: a short-lag category for Spanish /t/ and long-lag category for English /t/.

**Realization Rules**

It late learners are limited to merely approximating the VOT norm for English /p,t,k/ because they are unable to establish phonetic categories for these similar L2 stops, one might wonder how they manage to produce a VOT difference between the /p,t,k/ of Spanish and English. We hypothesize that they do so by applying different realization rules to a single phonetic category. According to the speech production model sketched in figure 1 (adapted from Keating 1984), phonologically voiced and voiceless stops are implemented by one of three universal phonetic categories corresponding to the traditional distinction between voiced, voiceless unaspirated, and voiceless aspirated. These categories specify broad articulatory (hence acoustic) characteristics of stop consonants. The greater detail needed to account for small but systematic cross-language differences in dimensions like VOT are supplied by language-specific realization rules. This production model might be used to account for bilingual production patterns. By hypothesis, Spanish late learners differentiate Spanish /t/ from English /t/ by applying two different realization rules to a phonetic category originally established to implement just the /t/ phoneme of Spanish. Early learners, on the other hand, produce a more substantial L1 versus L2 difference by implementing /t/ with different phonetic categories in Spanish and English, each motorically output by its own realization rule.

The production model just outlined was evaluated by having Spanish/English bilinguals produce Spanish and English /t/ initial words at the end of alternating Spanish and English sentences and phrases, and in isolation (Flege 1991c). It was predicted that the late

---

9In the first condition, the /t/-initial words occurred at the end of English and Spanish sentences “Take another word such as __” and “Tengo palabras como __”. In the second condition, the same words occurred at the end of English and Spanish phrases (“Take a ____”, “Tengo un ____”). The words were spoken in isolation in the third and final condition.
Vowel Production

A recent study of English vowel production in our laboratory (Flege and Bohn 1992) also points to important differences between early and late learners. We analyzed five tokens of English words (/beat, bit, bet, but/) spoken by ten subjects in each of five groups: two English monolingual groups (one from Alabama, one from Texas), two groups of Spanish late learners (one made up of individuals who had lived in the U.S. for 0.4 years on the average, the other 9.0 years), and a group of Spanish early learners. As shown in figure 2, both the early and late learners closely resembled native English speakers in producing temporal contrasts between English vowels. The talkers in each group made /i/ longer than /ɪ/, and /æ/ longer than /ɛ/. Figure 3 and figure 4 show that the native English speakers and the early L2 learners, respectively, produced these four English vowels with little spectral overlap. It is apparent from figure 5, however, that the late learners differed considerably from the native speakers of English. Both late learner groups produced a larger spectral contrast between /ɛ/-/æ/ than the native speakers. Neither the experienced nor the inexperienced late learners managed to produce a significant spectral contrast between /i/ and /ɪ/. These results support the hypothesis (Bohn and Flege 1990a, b...
that late learners will note temporal differences between a pair of unfamiliar L2 vowels more readily than spectral differences.

Flege and Bohn (1992) had three native English listeners identify vowels in the /bV/ words spoken by the native English and Spanish subjects using one of seven keywords (beat, bit, bet, bat, bat, but, bottle). Vowels spoken by the native speakers and the early learners were identified correctly almost without exception, but not vowels spoken by the late learners. The correct identification rates for the experienced late learners were /i/-57%, /I/-61%, /e/-99%, and /æ/-73%; those for the inexperienced late learners were /i/-69%, /I/-51%, /e/-91%, and /æ/-70%. For both groups, /i/ were heard as /i/, and vice versa. Given the traditional description of Spanish /i/ as being somewhat lower than English /i/ (see Flege 1988a), this suggests that the late learners may often have used Spanish /i/ to produce both English /i/ and /i/. The rates for /i/ were slightly higher for the experienced than inexperienced subjects, some of whose /æ/ attempts were heard as /i/. The inexperienced subjects' /æ/ was heard as /i/ in 30% of instances; the experienced subjects' /æ/ was heard as both /i/ (10%) and /i/ (13%).

These results clearly failed to support the hypothesis that the late learners would be able to master two English vowels we had considered new, viz. /i/ and /æ/. With the exception of /æ/, the late learners' vowels were identified at far lower rates than vowels spoken by native English speakers. The experienced late learners' /æ/ and /i/ attempts were not more identifiable than the inexperienced late learners' attempts. Despite these negative results, it would probably be unwise to reject the hypothesis that late learners are able to master new L2 vowels. Recall that support for the hypothesis was obtained in earlier studies (see above). Perhaps an L2 vowel will be treated as "new" only if it is found in a portion of the acoustic phonetic vowel space that is unoccupied by any allophone of an L1 vowel category. Such may be the case for English /æ/ but not English /i/, which is located in a portion of the space occupied by Spanish /i/ (especially the realizations of Spanish /i/ in closed syllables). Interestingly, we noted important differences between the three native English listeners, one of whom identified correctly nearly all of the late learners' /æ/ tokens. We are currently conducting additional listening tests in an attempt to determine why the late learners' /æ/ (and /i/) were correctly identified by certain listeners but not by others.
584 I Flege Speech Learning in a Second Language I 585

Another reason for not rejecting the "new vowel" hypothesis is the possibility that more native-like performance for /æ/ and /ɒ/ might have been obtained had other words containing these vowels been examined (perhaps words first encountered after category formation had taken place), or had we examined vowels spoken conversationally. Somewhat paradoxically, perhaps, there is evidence that L2 speech may be more authentic when it is produced in a spontaneous conversation than when it is read in a highly formal style, as is typical of phonetics experiments (see, e.g., Oyama 1976; Wenk 1985). Most important, however, is the evidence that some of the L2 learners seemed to have confirmed the prediction that new vowels can be mastered. As shown in figure 6, the correct identification rates for individual subjects varied considerably. Some late learners produced English /æ/, and even /ɒ/, quite well.

To obtain additional insight into the learnability of L2 vowels, we extended the English vowel intelligibility study to native German, Mandarin, and Korean late learners. Each L1 group consisted of 10 experienced and 10 inexperienced subjects. The difference between experienced and inexperienced subjects in the four L1 groups studied was generally quite small and usually nonsignificant. The mean values in figure 7 have therefore been based on all 20 subjects in each L1 group. The late learners' correct identification rates were considerably lower than the near-perfect rates of the native English speakers in nearly every instance. Exceptions to this general observation are the Germans' /i/ and /ɪ/ productions and the Spanish speakers' /e/ productions. Previous research for German (Bohn and Flege 1989) and Spanish (Dalbor 1980) suggest that these English vowels may have been so similar to vowels in the learners' L1 that, even if the L1 vowel were substituted for the corresponding L2 vowel without any modification, it would probably have gone unnoticed by the native English listeners.

An analysis of vowel confusions revealed differences between the four L1 groups that could be understood in terms of cross-language
Figure 6. The mean rate of correct identifications of the English vowels /a/, /i/ (top) and /e, /æ/ (bottom) as spoken by the native speakers of English from Birmingham, early L2 learners, experienced late L2 learners, and inexperienced late L2 learners (10 subjects per group). Each mean value is based on 75 forced-choice responses (3 listeners × 5 vowels × 5 replicate presentations).

Our review of the literature suggested that at least one of the four English vowels examined differed sufficiently from vowels in German, Mandarin, Spanish, and Korean to be considered "new." However, even though many of the L2 learners we examined had lived in the U.S. for many years and used English daily, the overall intelligibility rates for the English vowels were usually quite low in the absence of semantic context, some as low as 50%. Based on the acoustic data presented earlier, we thought the German subjects had mastered the /æ/ phoneme, but the intelligibility data suggested otherwise (see Bohn and Flege 1992a, for a discussion). The correct identification rates for the Korean subjects were quite low, especially for /æ/. Korean is not analyzed as having an /æ/, but a perceptual study using a matrix of synthetic vowels (Scholes 1968) indicated that nearly all steady-state, 400-msec vowels identified as /æ/ by native English speakers were consistently identified by Korean listeners in terms of a Korean vowel category rather than being judged as falling outside the Korean vowel inventory (i.e., labelled as not being an L1 vowel). Thus /æ/ might not have been treated as new by Korean learners of English as we originally thought it would be.

As shown in table 1, there were usually a few talkers in each L1 group who managed to produce English vowels at rates exceeding 90% correct. For example, nearly half of the Spanish L2 learners were successful in producing English /i/ and /æ/ at this level or better. This last finding shows that cross-language phonetic interference is not inevitable, for some of the late L2 learners were successful in producing an English vowel not found in their L1. There is, of course, no guarantee that this finding would generalize to other speaking styles and situations. It is not known what effect variations in speaking style may have had. As noted earlier, some of the subjects might have produced the English vowels more authentically in casual speech (see, e.g., Oyama 1976). It is also possible, of course, that list-reading tasks yield samples of L2 learners' optimal performance, and thus tend to overestimate their control of the L2 vowel system.
production seen in table I emerged and stabilized early in the process of L1 learning, perhaps during the first six months. Results obtained in a recent study (Flege 1992) provided preliminary evidence that prosodic errors, such as errors in stress placement or rhythm, may contribute less to FA than do segmental errors. Mueller and Niedzielski (1963) stated that FA derives principally from mismanagement of "allophonic details" in similar sounds. If so, then much of the improvement in FA that does occur in early stages of L2 learning may be due to an improved production of "new" sounds.

We are currently unable to account for the striking individual subject differences in vowel production that have been observed. Several kinds of explanation come to mind. Lambert (1977) provided evidence that some L2 learners ("code users") are more likely to perceive an L2 sound that differs auditorily from any L1 sound in terms of an L1 category whereas others ("code formers") tend to develop new central representations in such instances. Mack (1988) discussed the hypothesis that some L2 learners may make greater use of "bottom-up" processing than do native speakers in attempting to comprehend an L2. L2 learners who show this tendency to the greatest degree may be especially good at L2 segmental production and perception. Extensive use of high-order information might cause L2 learners to overlook the actual substance of specific sounds in the L2.

Mochizuki (1981) and Mack, Tierney, and Boyle (1990) provided evidence suggesting that non-natives may remember English words more poorly than native speakers. This may be due in part to a nonoptimal match between incoming English phones and the non-natives' central phonetic representations. Flege, Munro, and Fox (1992) had Spanish and English listeners judge the degree of dissimilarity of pairs of vowels. One finding of this study suggested that L2 learners with good pronunciation of English were better able to perceive sounds at a phonetic level than were L2 learners who pronounced English poorly. Another finding was that Spanish subjects who pronounced English well judged pairs of English vowels to be more dissimilar than subjects who pronounced English poorly. This suggested an auditory basis for L2 speech learning ability, perhaps a difference in ability to store and access sensory information for unfamiliar L2 sounds (see also Mayberry and Fischer 1989, for parallels in the learning of ASL).

A CRITICAL PERIOD?

FA is widely believed to be the consequence of a diminished ability to learn speech (Lenneberg 1967). The deficit is usually understood to be the result of faulty speech production, which implies a breakdown in
sensorimotor learning ability (Sapon 1952). Scovel (1988) claimed that humans cannot learn to speak an L2 without accent after about the age of 12 years because of lost “plasticity” in “neuropsychological mechanisms” that derive from the completion of cerebral lateralization. Because errors in segmental articulation contribute importantly to the perception of FA (e.g., Flege 1992), the data presented earlier, which showed more authentic production of L2 sounds by early learners than by late learners, could be viewed as supporting the existence of a critical (or sensitive) period for speech learning. It may be unwise to draw this conclusion, however (see Flege 1987a) because a clear relationship between L2 learning ability and lateralization has never been established. Also, chronological (and developmental) age is confounded with so many other factors that it would be difficult, if not impossible, to demonstrate convincingly the existence of a critical period (see below). Even if a critical period were shown to exist, it would provide little insight into why adults speak with a FA.

Also, recent experiments have confirmed earlier evidence (e.g., Asher and Garcia 1969) that FAs emerge before the age of 12–14 years, as commonly supposed (Patkowski 1989). Flege (1988c) found that native Chinese adults who began learning English at an average age of 7.6 years spoke English with a slight but detectable accent. A later study using the same stimuli and procedures showed that native Spanish adults who learned English by the age of five to six years did not differ significantly from native English speakers (Flege and Fletcher 1992). Although further research is clearly needed, the two studies suggest that FA first becomes evident at some time between the ages of five and seven years.

There are potential explanations for the emergence of an FA other than the passing of a critical period that should be considered. For example, an FA might arise in adults and older children because they do not receive L2 phonetic input that is as rich as the phonetic input received by young children learning an L1 (see Flege 1987a). Adults want and need to communicate in an L2 as soon as possible, so they may begin speaking before they establish accurate perceptual representations for L2 sounds. This could result in incorrect production patterns that are difficult to change later, when more accurate perceptual representations may have been developed (Henning 1966). Alternatively, L2 learners may be less likely than L1 learners to form accurate perceptual representations for L2 sounds. This would probably be the case if the auditory processing of L2 words terminates earlier in adults and older children as the result of more rapid word recognition.12

---

More rapid word recognition by adults than young children might come about through the greater, or earlier, use of higher order information in parallel with bottom-up phonetic information.

---

A classic view of perceptual development is that children become increasingly less reliant on sensory information as they develop cognitively and learn to ignore attributes of sensory stimuli that are irrelevant to classification. Age-related changes in psychosocial or socioaffective factors may account for differences between older and younger L2 learners (e.g., Schumann 1975; Krashen 1985). Others have pointed to cognitive changes that may affect how L2 input is operated upon (Schneiderman and Desmarais 1988). However, a cognitive approach has been applied less often to L2 speech than morphosyntactic learning because of the sensorimotor component of speech production and perception.

CHANGES IN SPEECH LEARNING

The general hypothesis that has motivated our program of speech research over the past decade (e.g., Flege 1981) has been that perceptual and sensorimotor processes that permit children to learn to pronounce their L1 without an accent remain intact through the lifespan, that is, do not deteriorate (or become inaccessible) as the result of neurological maturation. We have hypothesized that these processes operate in much the same way in L2 and L1 acquisition, but may yield different results in L2 learning than in L1 acquisition because the phonetic system has often stabilized by the time L2 learning begins. Early learners may learn to pronounce an L2 better than late L2 learners because the late learners are first exposed to the L2 after an important developmental shift in speech processing that has occurred at around the age of five to seven years. Recall that this appears to be the age at which foreign accents first emerge (Flege and Fletcher 1992).

We propose that an important consequence of the hypothesized “phonetic system shift” is that it makes L2 learners more likely to equate L2 sounds with sounds in the L1. This may render late learners less able to establish additional phonetic categories for sounds in the L2 after rather than before the age of five to seven years. If so, it may mean that foreign accents do not arise because speech learning ability has diminished as the result of neurological maturation (Scovel 1988), but because the phonetic system has developed and stabilized. Treiman and Zukowski (1990) presented evidence that the smallest portion of a word that might be used in performing explicit perceptual tasks by children between the ages of four and six years changes from a whole syllable, to part of a syllable (i.e., the onset or rime), to a single segment. Valley (1990) suggested that such developmental changes may reflect how words are represented in the lexicon (see also Fowler 1990). Children’s
first words may be represented holistically whereas words added to the lexicon in later childhood may be stored and retrieved in terms of phonemic segments (Ferguson and Farwell 1975; Ferguson 1986; Jusczyk 1986).

Development of the phonetic system may, itself, be the indirect result of general cognitive changes involving how abstract concepts are formed. The ability to perform tasks that focus on specific formal aspects of a word (e.g., to say the word pat without its first sound, or to find the commonality in word pairs such as player-prayer, cat-cad) requires explicit, metalinguistic awareness. Cognitive psychologists speak of a “5 to 7 shift” reflecting basic changes in learning strategies and concept formation (e.g., White 1965; Shaffer 1985). The ability to focus attention on specific aspects of stimuli, called decentering, epitomizes the stage of concrete operations said to begin between the ages of five and seven years (Piaget and Inhelder 1969). Decentering allows the child to consider alternative hypotheses in attempting to solve a problem. One study showed that five-year-olds who performed poorly on a measure of concrete operations were less successful in reading a year later than children who performed well on the concrete operations measure (Tunmer 1988).

Reading, of course, requires phonemic awareness. Nettouer and Studdert-Kennedy (1987) found that, in identifying fricatives, children aged three to five years were more influenced by information in the following vowel than seven-year-olds and adults, who tended to base their judgments on spectral properties of the friction noise itself. It was as if the youngest children were unable to focus attention on just the fricatives. An increased awareness of segment-sized units in speech processing appears to be motivated by lexical growth. As the number of lexical items increases, so too does the likelihood that “near neighbors” in the lexicon will be confused with one another if the sounds which differentiate them are not represented explicitly (Jusczyk 1989; Charles-Luce and Luce 1990).

The finding that children show increasing perceptual awareness of segment-sized units agrees with the finding that young children may show greater coarticulation than adults in their production of speech as the result of producing words in a somewhat less “segmental” fashion than adults. In a study comparing adults to children aged three to seven years, Nettouer, Studdert-Kennedy, and McGowan (1989) found an increase with age of the spectral contrast between /s/ and /ʃ/ fricatives. At the same time, they noted a decrease with age in the effect of vowel context on the fricative spectra. This suggested that, unlike adults and older children, young children's word production may be organized over a domain “at least the size of the syllable,” only later reflecting an organization that is “aligned with perceived segmental components” (Nettouer, Studdert-Kennedy, and McGowan 1989, p. 120).13

There are many such instances of whole-to-part shifts in perceptual development (Aslin and Smith 1988). An increased awareness of segments probably does not occur abruptly; and its timing may differ across sound types. Nor is it likely that increased segmental awareness, in itself, will account for the emergence of foreign accent. We propose that as children’s awareness of segments increases during the period from five to seven years, the categories comprising their phonetic system undergo two broad types of change: (1) the core acoustic properties of prototypical exemplars of each phonetic category, and the weighting of those properties, will become better defined; (2) the range of phones that are identifiable as a realization of each category will increase. Note that the first hypothesis pertains to category centers, the second to the boundaries between categories.

These two hypotheses concerning phonetic system development are illustrated in figure 8, which shows four hypothetical vowel categories at two times, one prior to the hypothesized phonetic system shift (Time 1), the other after the shift (Time 2). Each box is meant to illustrate a portion of the phonetic vowel space. The two dimensions of the boxes might be thought of as representing acoustic properties such as F1 and F2 frequency but, more realistically, the vowel space should be thought of as an N-dimension psychological space. Four vowel categories, which are represented by circles, change in several respects from Time 1 to Time 2. The categories embrace a wider range of variants (indicated by the increased size of the ellipses); have better defined centers, or prototypes (illustrated by the darkened circles at the approximate center of each category); and have better defined boundaries (indicated by the thickness of the lines). From the standpoint of L2 learning, the most important feature of the change from Time 1 to Time 2 is the reduction of “uncommitted” vowel space not occupied by any L1 vowel category. The hypothesis being advanced here is that unless most realizations of an L2 vowel are found within “uncommitted” space, the L2 vowel will tend to be identified in terms of an L1 category so that new category formation will not occur. Thus, an L2 learner

13It is worth noting that segmentation skills may influence many aspects of speech processing, not just L2 learning. Good readers show better segmentation skills than poor readers. Mann (1984) concluded that good readers use phonetic representations more effectively to process speech than poor readers. Burnham (1986) found that children with good comprehension abilities for their age were more likely to identify sounds in accordance with the phonemic categories of their L1, and to ignore phonetic contrasts that were not phonemically relevant in L1, than children with relatively poor comprehension abilities.
should be more likely to add new L2 vowel categories at Time 2 in phonetic development than at Time 1.

Based on children's production and perception of vowel duration cues to the contrast between voiced and voiceless stops in word-final position, Krause (1982a, b) concluded that both speech production and perception demonstrate "refinement and stabilization" with increasing age (1982a, p. 25). A better definition of phonetic category prototypes may better enable children to gauge degree of FA as they grow older (Scovel 1988) and to gauge how "clear" or "distorted" the realizations of a particular category may be. The outward expansion of category boundaries should enable children, as they grow older, to correctly identify sounds produced by an ever wider variety of talkers differing in vocal tract size, dialect, and characteristic speaking rate.

Neither of these hypotheses have as yet been tested formally, but they seem to be consistent with previous results. Based on experiments examining the identification of the voicing feature in stops, Simon and Fourcin (1978) concluded that response variability decreases over much of childhood, and a "categorical" mode of response increases up to about the age of five to seven years (depending on the children's L1). This is what one would expect if phonetic category prototypes became better defined. Zlatin and Koenigsknecht (1975, 1976) also examined the identification of stop consonants by children. They found that the slopes of young children's identification functions for voice onset time (VOT) continua were not steep. This meant that a larger number of stimuli from the VOT continua were judged consistently as belonging to either the voiced or voiceless stop category by adults and older children than by young children.

The decrease with age in the number of "ambiguous stimuli" in a forced-choice task implies that the two categories defined by the endpoints of a perceptual continuum embrace a wider range of phones. The hypothesis that the range of phones identifiable in terms of LI categories expands with age received indirect support from an experiment by Butcher (1976). His experiment assessed the degree of perceived similarity of cardinal vowels by adults and children who spoke English, French, or German. Cross-language comparisons indicated that cardinal vowels were perceived to be closer when they occupied crowded than uncrowded portions of the listeners' LI vowel space. Perceived dissimilarity ratings (which presumably reflected intervowel distances in a psychological vowel space) tended to be less for adults than children. This suggested that the adult's phonetic space was "filled" to a greater extent than the children's.

LINGUISTIC FACTORS

This chapter has dealt with how sounds in an L2 are produced and perceived. From a linguistic (i.e., phonological) perspective, the research reported here might fall under the general heading of "segmental structure constraints." Although we have focused on segments, many other factors that fall outside the scope of this chapter may influence L2 pronunciation. In particular, properties of language production that encompass more than a single segment are likely to influence segmental production and perception. For example, in agreement with the observation that "segmental phonemes are influenced by the suprasegmentals of language" (Mueller and Niedzelski 1963), Wenk (1985) found that vowel production accuracy may depend on L2 rhythmic accuracy.

Just as L2 segments may or may not be identified with segments in the L1, learners may (or may not) identify strings of segments in the L2 as meeting the structural description for the application of an LI phonological rule or constraint (Broselow 1983; Rubach 1984; Major 1987c;
For example, Hammarberg (1990) refers to a beginning German learner of Swedish who pronounced the Swedish word for “bed” ("säng") with an initial /s/ when prompted, but with an initial /z/ (typical for German) once he had learned the word and could produce it spontaneously. The phonological principal of underspecification may be needed to account for which vowel from the L1 inventory is inserted into L2 words in order to make them conform to L1 syllable structure constraints (Weinberger 1990). It might also be needed to account for instances of “differential substitution” (e.g., the use of /s/-for-/z/ by Japanese learners of English but /t/-for-/l/ substitutions by Russians, even though both L1s have /s/ and /l/).

It is likely, too, that lexical factors will have an important impact on L2 segmental articulation and learning. Mack, Tierney, and Boyle (1990) found that both native speakers and L2 learners recognized relatively frequent (and presumably familiar) words better than infrequent ones, especially when the words were generated by a computer. Conversely, L2 learners may produce segments more authentically in relatively infrequent than in frequent words. This might be expected if L2 learners tend to mispronounce sounds in the first L2 words they learn because phonetic categories for the L2 sounds have not yet been established, then have difficulty later in correcting these mispronunciations once categories have been established for the L2 sounds.

L2 words that have a cognate in the L1 may be produced differently than words that cannot be related to a known L1 word. Wenk (1985) observed more accurate rhythmic properties in L2 words that had a close cognate in the L1 than in those without a cognate. In a recent study, Flege and Munro (1992) found that VOT values in English /t/ tokens were less authentic in an English word with a Spanish cognate (taco) than in noncognate words. Differences between cognate and noncognate words may be especially important when L2 learning is mediated by the written word. Hammarberg (1990) found that Germans with no knowledge of Swedish imitated Swedish /a:/ (a low, back slightly rounded vowel) with an [ø]-quality vowel whereas Germans who knew some Swedish produced it using an German-like [ø]-quality vowel. Orthographic cues may encourage German learners of Swedish to use a “top-down” solution (whereby the Swedish /a:/ is identified with German /ø/) rather than a more phonetic “bottom-up” solution. The use of LI vowels was especially apparent in Swedish words that had lexical equivalents in German (e.g., radió).

SUMMARY

It seems likely that how well an L2 sound that differs acoustically from any sound in the L1 is produced will depend, to a large extent, on how it is categorized. It is proposed here that the primary difference between early and late L2 learners is that late learners are more likely than early learners to equate L2 sounds perceptually with sounds in the L1, which may prevent late learners from establishing phonetic categories for certain L2 sounds and thus producing them authentically. Another proposal is that the emergence of foreign accent in an L2 at about the age of five to seven years results from the decreasing likelihood that L2 sounds will be judged as falling outside of the existing L1 inventory. This important shift may coincide with an increase in children's segmental awareness, and be linked to changes in the structure and content of their phonetic categories. Paradoxically, the changes in phonetic categories probably result in increased efficiency in perceiving and producing LI sounds, but they may make it increasingly difficult for children to master sounds not found in the LI phonetic inventory. If we are correct in thinking that many L2 production errors have a perceptual basis, this may explain why so many skills, including speech motor control (e.g., Smith 1978), seem to increase with age whereas the ability to learn new forms of pronunciation seems to decline.

REFERENCES


Bohn, O.-S., and Flege, J. E. 1990a. Interlingual identification and the role of...


Ferguson, C. 1990. Personal communication.


Mack, M., Tierney, J., and Boyle, M. 1990. The intelligibility of natural and


