Chapter 5

The intelligibility of English vowels spoken by British and Dutch talkers

James Emil Flege

University of Alabama at Birmingham

1. Introduction

Intelligibility testing first began as a means for evaluating the effectiveness of communication devices such as telephones (Fletcher and Steinberg 1929). It was later used as a way to assess the articulatory effectiveness of talkers (Black, 1976) and speech synthesizers (Green, Logan, and Pisoni, 1986; Logan, Green and Pisoni, 1989). In more recent years, the role of the listener in assuring the effective transmission of messages via speech has been recognized (e.g. Rubin, 1983; Pisoni, Nusbaum and Green, 1985). The chapter presents a model of second language (hereafter, L2) speech learning. The intelligibility of vowels spoken in a second (foreign) language is then examined. The results show that the phonetic relationship between vowels in the L2 and in the native language (L1) is an important determinant of the intelligibility of L2 vowels. It is also shown that vowels differ inherently in intelligibility, and that the native dialect of the listener plays a role in determining the intelligibility of vowels spoken in an L2.

1.1 Establishing a phonetic inventory

Young children just beginning to learn speech must discover which classes of phones (hereafter, sounds) in their L1 are contrastive. They must also learn how to articulate sounds according to the phonetic norms of their L1. For example, a native French child must learn to produce /t/ as a voiceless
unaspirated stop with a dental place of tongue contact ([t]) whereas a native English child must learn to produce /t/ as a voiceless *aspirated* stop with an alveolar place of constriction ([th]).

Mastery of many sounds requires a period of skill acquisition. Although children may misarticulate certain L1 sounds for a time, most sounds are eventually mastered once learning has run its full course. One important characteristic of L1 acquisition is so obvious that it is seldom noticed: although they may misarticulate their L1, children never seem to speak it with a foreign accent. This is because foreign accent does not derive primarily from insufficient learning, but from the inappropriate use of sounds developed for one language in speaking another.

When attempting a sound that does not occur in the L1, L2 learners — at least relatively inexperienced ones — often substitute the nearest L1 sound. For example, a Spanish speaker is likely to produce the English word *beat* with a Spanish /i/. Spanish and English both have an /i/, but Spanish /i/ is produced with a slightly lower tongue position than English /i/ (Flege, 1989b). Many Spanish speakers of English, even those who are highly experienced in English, seem to use a Spanish /i/ in English words (Flege, 1991b). To take another example, a German may produce English *bat* with an [ɛ]-quality vowel because German has no /æ/ (Bohn and Flege, 1991).

Nearly all L2 speech errors involve sounds which either do not exist in the L1, or differ from their L1 counterpart (James, 1984, 1985a). For example, Flege and Hillenbrand (1984) found that adult native French speakers of English produced English /t/ with voice onset time (VOT) values that were intermediate to the phonetic norms of French and English for voiceless stops, defined in part by the mean short-lag VOT values measured in /t/s produced by French monolinguals and the long-lag VOT values observed for English monolinguals. Stops with “compromise” VOT values were also observed in L2 stops produced by adult native speakers of English who had learned French.

The nonauthentic production of L2 /t/s by the French and English subjects in the Flege and Hillenbrand (1984) study was probably *not* the result of an inability to detect auditorily the phonetic difference in how /t/ is implemented in French and English (Flege, 1984; see also Flege and Ham mond, 1982). The L2 learners’ ability to note at least some of the acoustic differences that distinguish the /t/s of French and English was demonstrated by the fact that both French and English subjects were able to modify /t/
partially when speaking their L2. This raises the question: Why did the L2 learners not modify their production sufficiently to permit an *authentic* production of the L2 /t/?

1.2 *The Sensitive period Hypothesis*

One explanation offered frequently for the unmodified substitution of L1 for L2 sounds, or for the partial modification just mentioned, is that a "sensitive period" exists for speech learning (see Snow, 1988 and Flege, 1987b). In a recent cross-sectional study examining the effect of age on L2 learning, Johnson and Newport (1989) found that second language learners’ ability to learn L2 syntax begins to diminish gradually (or is used less) long before puberty. If the same phenomenon applies to speech learning, then one might suppose that young children learn to pronounce their L1 without accent because L1 learning occurs during early childhood, a time when speech learning ability is presumably at its height. By hypothesis, learning the sound system of an L2 after the sensitive period has passed will be less rapid and/or complete because speech learning ability has diminished or for some reason is only partially exploited.

The French and English subjects examined by Flege and Hillenbrand (1984) were living in an environment where L2 was the predominant language. They had used the L2 as their primary language for an average of 12 years, and appeared to be highly motivated to pronounce the L2 authentically. It is therefore unlikely that these subjects would ever approximate the phonetic norms of their L2 for /t/ more closely than they already had. This suggests that an upward limit exists for the learning of certain L2 sounds, and leads to the following sensitive period hypothesis:

H1: Adult L2 learners are less able than child L2 learners to translate into gesture, via the establishment of central sensorimotor representations, the auditory and visual stimulation that accompanies the production of L2 sounds not found in the L1.

By hypothesis, the French subjects examined by Flege and Hillenbrand (1984) were less able than young children learning English as an L1 to learn the pattern of glottal-supraglottal timing needed to implement /t/ as [tʰ]. If one assumes that producing stops with short-lag VOT is somehow basic or unlearned (e.g., Kewley-Port and Preston, 1974) and that all new learning is mediated by previous learning, then hypothesis could be reformulated as follows:
H2: Adult L2 learners are less able than young children to modify existing (L1) pattern of speech production in order to produce L2 sounds authentically.

A sensitive period hypothesis is consistent with the belief that individuals who learn an L2 as young children often speak both of their languages without a foreign accent, whereas those who learn an L2 later in life typically speak it with a foreign accent (see Tahta, Wood and Loewenthal, 1981). For example, Flege (1988a) found that native English-speaking listeners gave significantly lower (i.e., more foreign accented) scores to English sentences spoken by Chinese subjects who learned English as adults than to sentences spoken by native speakers of English. Many have supposed that the offset of a sensitive period for speech occurs around the time of puberty, but another finding of the Flege (1988a) study suggested an earlier offset: Chinese subjects who began learning English L2 at an average age of eight years also produced English sentences with a detectable foreign accent.

The perception of foreign accent is based on many aspects of segmental and suprasegmental articulation. Acoustic analyses of speech sound production have also been consistent with a sensitive period hypothesis. Flege (1991) found that native Spanish children who began learning English as an L2 in a Texas school at about the age of 5-6 years of age were able to produce both English and Spanish /t/s authentically. These “Early L2 Learners” produced Spanish /t/ with appropriate short-lag VOT values, and English /t/ with appropriate long-lag VOT values. “Late L2 Learners” who began learning English L2 in adulthood, on the other hand, were unable to differentiate fully Spanish and English /t/. As in many previous L2 production studies, the Late L2 Learners produced English /t/ with “compromise” VOT values that were intermediate to the phonetic norms of Spanish and English.

2. Purpose of the chapter

A sensitive period hypothesis seems to account for the well-known phenomenon of foreign accent, but there are several problems with it, at least as just formulated. For example, a sensitive period hypothesis provides no explanation for why the French speakers of English examined by Flege and Hillenbrand (1984) only partially approximated the English phonetic norm...
L2 VOWEL INTELLIGIBILITY

for /t/. Why should it be any easier to modify laryngeal timing to give the approximately 20-ms increase in VOT that was observed than to learn a modification that would give the approximately 40-ms increase needed to achieve the long-lag phonetic norm of English? The same thing can be asked of the native English learners, whose learning task was to shorten VOT, that is, to produce French /t/ with short-lag VOT rather than the accustomed long-lag VOT values.¹ Nor does a sensitive period hypothesis offer any insight into what specific speech learning mechanisms or processes are changed or attenuated as humans mature physiologically and develop cognitively.

The speech production data now available make it appear likely that previous L1 learning affects subsequent L2 learning through the intermediary of central cognitive-linguistic and phonetic structures more abstract than the sensorimotor level implied by a sensitive period hypothesis. Therefore, the present chapter has two major purposes: (1) to definiate a speech learning model (henceforth SLM) that attempts to describe the mechanisms and processes by which L1 phonetic interference affects the production of L2 vowels and consonants and account for age-related differences in the learning of sounds in an L2; and (2) to present the results of a vowel production study that provided a preliminary test of the model.

Most empirical L2 research has focussed on consonants, especially the VOT dimension in word-initial stops. The study presented in this chapter will test the SLM by examining the production of six English vowels (/i, æ, a, u, I, A/) spoken by 50 Dutch university students who began learning English in school at about the age of 12 years. Flege and Eefting (1987) previously obtained assessments of the degree of foreign accent in English sentences spoken by the Dutch subjects. As predicted by a sensitive period hypothesis, all 50 had detectable foreign accents even though many were majoring in English and had frequent access to native speakers of English at school or during trips to England. The Dutch subjects’ success in learning English vowels was assessed by determining how often each of the six vowels was identified by native English-speaking listeners.

As noted by Leather (1983: 210), one general problem in testing speech learning hypotheses is that auditory perceptual judgments, even those of trained observers, are likely to be unacceptably inconsistent unless measures are taken to reduce measurement error. Thus some of the many factors (other than articulation) that can influence vowel intelligibility will
be considered briefly below before turning to specific predictions of the SLM.

Development of the SLM was initially prompted by the observation that, in general, sensorimotor skills increase through childhood and the adolescent years. A number of studies have shown that the ability to imitate sounds not found in the L1 increases with age (Politzer and Weiss, 1969; Olson and Samuels, 1973; Snow and Hoefnagel-Höhle, 1982a; Ekstrand 1982). Why then, should speech learning ability decline with age? The SLM attempts to account for the psycholinguistic and phonetic processes that underlie speech learning. The SLM is in accord with the general principle that sensitive periods are generally centered around the period when learning takes place most rapidly, and that stable systems are less susceptible to change than rapidly evolving ones (Bornstein, 1987).

One important hypothesis generated by the SLM is that the phonetic categories needed to produce and perceive L2 sounds rapidly and accurately in conversational speech can be established readily until about the age of 5-6 years, when the phonetic system begins to stabilize. After that age, additional categories can be established for “new” but not “similar” L2 sounds. Although both new and similar (but not “identical”) L2 sounds differ acoustically from sounds in the L2, there is thought to be a qualitative as well as a quantitative difference in the degree of phonetic dissimilarity between new and similar sounds and sounds found in the L1. A new L2 sound is defined as an L2 sound that differs sufficiently from any sound in L1 that it evades the effects of equivalence classification, a basic cognitive mechanism thought to shape both L1 and L2 speech learning. The distinction between new and similar sounds, which remains elusive, will be described further below.

The SLM assumes a continuity in the mechanisms used to acquire the L1 and to learn an L2 later in life. It posits that speech learning does not necessarily end when the L1 has been “mastered” because humans continue to learn phonetically whenever they are required to communicate via a phonetic system that differs systematically from the one(s) they have used previously. More specifically, the SLM hypothesizes that the phonetic system may evolve over the lifespan to permit the addition of phonetic categories for new L2 sounds that occupy a portion of the phonetic space that has not been exploited previously; and the model posits the updating of any existing phonetic category so that it may better reflect the acoustic substance of the wide array of phones in L1 and L2 that have been identified as
realizations of it. The SLM also hypothesizes that changes in motoric processes are possible following perceptual modifications: additional phonetic realization rules may be developed to output new and modified phonetic categories.

The SLM, as formulated below, leads to the prediction that the Dutch subjects would be more successful in learning new than similar English vowels. However, before predictions generated by the SLM are discussed, assumptions about the nature of vowel production and perception will first be presented and the current procedures used to classify L2 vowels as similar and new (or as “identical”) described.

2.1 Speech production and perception

Williams and Nottebohm (1985) identified direct links between motor output and auditory input for male zebra finches, an avian species which completes song learning about 90 days after hatching and shows no changes in song thereafter, even after deafening. Echoing the “motor theory” of speech perception, the authors concluded that conspecific sounds may be decoded “by reference to the vocal gestures used to produce them” (p. 279). Unfortunately, the linkage between vocal production and perception in less well understood in humans than in birds. It is uncertain how best to account for the observation that, in infancy, the effect of the surrounding linguistic environment begins to influence perception before it influences production (Jusczyk, 1989), or why young children are often able to comprehend words more accurately than they can produce them. Divergences in productive and perceptual abilities has suggested to some the existence of separate lexicons for production and perception (Menn, 1982), and to others (e.g., Matthei, 1989) a single lexicon with separate productive and perceptual access routes to abstract linguistic information.

The SLM hypothesizes that speech production and perception are both organized at three sequential levels. A consideration of the speech of adult monolinguals suggests that a broad correspondence exists between production and perception at the three levels, especially at the most abstract, phonemic, level. However, the structures that regulate the encoding and decoding of the phonic elements needed for aural-oral communication may not be isomorphic in adult L2 learners. This was shown recently by the finding that Chinese speakers who themselves produced English with a foreign accent were as able as native English speakers to differentiate native from
Chinese-produced sentences (Flege, 1988a). The assumption is made, as is common in both L1 and L2 research (e.g., Gottfried and Beddor, 1988: 72), that changes in perception are generally a necessary but not a sufficient condition for corresponding changes in speech production.

2.1.1 Speech production

Speech production is conceptualized as being organized at three levels: a phonemic level, a somewhat less abstract phonetic category level, and a sensorimotor level. As described by Keating (1984), phonemes are implemented by a finite set of universal phonetic categories which, in turn, are physically output through learned language-specific realization rules. Relatively little empirical research has focussed on the process of phonetic category formation or the evolution of phonetic categories. This is surprising because phonetic categories represent an essential interface between the abstract phonemic level, at which words are specified in the mental lexicon, and the sensorimotor representations used to articulate speech.

Results obtained by Flege and Eefting (1988) suggested that Spanish-English bilinguals may have two phonetic categories for implementing the phoneme /t/, one for producing voiceless unaspirated stop [t]s in Spanish and one for producing voiceless aspirated [th]s in English. The subjects imitated a VOT continuum ranging from /da/ to /ta/ that contained stops with lead, short-lag, and long-lag VOT values. The Spanish and English monolingual subjects produced stops with VOT values in only two of the three modal VOT ranges whereas native Spanish subjects who had learned English by the age of 5-6 years (Early L2 Learners) produced stops with VOT values in all three modal ranges.

In a previous study (Flege and Eefting, 1987b) these same subjects had identified the members of the VOT continuum in a two-alternative forced-choice test. The phoneme boundaries derived from that study coincided with the location along the continuum where VOT values in the vocal imitation responses shifted between modal VOT categories (i.e., from lead to short-lag, and from short-lag to long-lag). The identification data suggested that both the monolingual and bilingual subjects possessed two phonological categories (viz. /t/ and /d/), whereas the imitation data suggested that the bilinguals had three phonetic categories. One might conclude from these results that Early L2 Learners possess an enriched phonetic system containing phonetic categories from both the L1 and the L2, which may explain why early childhood bilinguals seem to produce both of their languages

Studies examining the L2 production of Late L2 Learners have shown that they may produce /p,t,k/ in the L2 with "compromise" VOT values. Flege (1991a) found, for example, that Spanish Late L2 Learners produced /t/ with significantly longer VOT values in English than Spanish, but nevertheless differed significantly from native English speakers. By hypothesis (see below), the Late L2 Learners had not established a phonetic category for English /t/; they succeeded in partially differentiating /t/ in Spanish and English through the use of a phonetic realization rule.

Realization rules specify the timing, amplitude, and duration of muscle contractions that position the speech articulators in space and time (e.g. Chomsky and Halle, 1968; Lieberman, 1970; Ladefoged, 1975, 1980, 1983). The need for realization rules in a model of speech production is suggested by the existence of small but systematic acoustic differences that may distinguish corresponding sounds in two languages. For example, both Danish and English have phonemes implemented as voiceless aspirated stops. By hypothesis, the slightly longer VOT values observed for Danish than English /p,t,k/ (Christensen, 1984) result from the application of different language-specific realization rules of course, such differences might be accounted for by the existence of different language-specific (rather than universal) phonetic categories. However, an independent justification for realization rules is the need to account for differences in segmental articulation that result from rate changes, requests for clarification, or for changes in interlocutor. For example, Labov (1981) noted that monolingual native speakers of English systematically altered their production of English /u/ by fronting this vowel in certain speaking situations.

An important determinant of intelligibility for speech produced by text-to-speech systems is the number and sophistication of realization rules which specify the parametric input to synthesis rules (Logan et al., 1989). In a way that parallels the development of speech perception and the refinement of text-to-speech systems, there is evidence that children learning their L1 add and/or refine realization rules gradually after discovering more basic means for implementing stop phonemes (e.g., Macken and Barton, 1980; Mack and Lieberman, 1985).

Within several years after the appearance of language, children can usually produce most sounds of their L1 recognizably. This implies that they have established phonetic categories with which to implement the
phonemes of their L1. Studies have shown that speech motor control continues to be refined for several years (e.g., Eguchi and Hirsh, 1969; Smith 1977). Other studies have shown that children’s production changes gradually so that their speech conforms ever more closely to the phonetic norms of their L1 (e.g., Zlatin and Koenigsknecht, 1976; Mack and Lieberman, 1985; Flege, McCutcheon, and Smith, 1987; Flege, 1988c), probably as the result of changes in realization rules. It is assumed here that adults remain able to add or refine phonetic realization rules.

2.1.2 Speech perception

The perceptual processing of speech is conceptualized as occurring at three levels of organization: phonemic, phonetic, and auditory.

The auditory level of analysis is earliest both in terms of ontogeny and on-line processing. Infants are able to discriminate most, if not all, of the phonetic contrasts exploited by human languages. They do so on the basis of innate auditory abilities in much the same way that macaques, for example, can discriminate between stops differing in place of articulation (Kuhl and Paden, 1983) or vervets can discriminate vowels (Sinnott, 1989).

Most but not all (e.g. Klatt, 1979) models of speech perception posit that words are recognized through the intermediary of phonetic and/or phonemic category representations. It appears that infants’ early sensory-based ability to discriminate speech sounds is modified once they begin to process speech sounds phonetically. For example, Werker and Tees (1982, 1983, 1984a,b) found that 6 to 8-month-old infants being raised in an English-speaking environment could discriminate a phonetic contrast that occurs in Hindi but not English (viz. a contrast between dental and retroflex /t/) whereas infants aged 10 to 12 months could not.3

The maintenance of an ability to discriminate some but not all of a “universal” set of potential phonetic contrasts (Werker, 1986; see also Jusczyk, 1986) suggests that infants soon acquire at least some knowledge of the segmental phonetic inventory of the language spoken around them. A recent study using a headturn preference procedure (Jusczyk, Friederici and Wessels, 1989) supports this view. Nine-month old infants from an English environment looked longer at a loudspeaker emitting English words than at another loudspeaker emitting Dutch words. The preference for words from the ambient language was not evident in six month-old infants, nor was it evident in the nine-month-old infants when the segmental information was removed by low-pass filtering.
Werker (1986) suggested that infants will remain sensitive only to those phonetic differences that have “phonemic significance” in the ambient language after their auditory processing of speech is “reorganized” at around 12 months of age. This reorganization, which Werker and others regard as stemming from reversible attentional processes, depends on infants’ ability to note the phonetic similarity present in a class of acoustically different phones. Kuhl (1979) found that infants could ignore nondistinctive acoustic differences in vowels such as pitch contour and information pertaining to the talker’s gender. Hillenbrand (1983, 1984) showed that infants could “sort” consonants into phonetically relevant categories.

Near the end of the first year of life, then, it appears that infants begin to organize disparate sound elements found in the ambient language into what might be termed phonetically-relevant categories. Data reported by Greiser (1984; Greiser and Kuhl, 1989) suggested that infants and native-English adults may possess similar prototypes for /i/. This raises the possibility that certain early-appearing phonetic categories may be “natural” categories, like focal colors (Rosch, 1975; Bornstein, 1975). However, given the wide range of sounds that occur in human languages, at least some phonetic categories are likely to be influenced by experience with specific varieties of speech.

Aslin and Pisoni (1980) hypothesized that children may “tune” innate auditory processes as the result of language-specific experience. Jusczyk (1989) suggested that attentional allocation is modified by learning, which may alter the weight given to various acoustic properties in the speech signal. The resulting “interpretative schemes”, which may develop through “innately guided” learning processes, are optimally suited for the rapid processing of phones found in the L1.

In most instances, listeners consciously perceive phones in terms of phonemic categories established as the result of previous perceptual learning (Gibson, 1969). English-speaking listeners, for example, ordinarily disregard the acoustic difference between the realizations of /t/ in words like tab and bat because they have learned that the audible acoustic differences between /t/s that are released and aspirated as opposed to those which are unaspirated and (often) unreleased are not used to signal differences in meaning in English. Although the [tʰ] and [t] allophones of /t/ are phonetically different, they are perceived to be the “same” at the most abstract — and conscious — level of categorization. Another way of stating this is to say that native speakers of English possess phonological knowledge that English words cannot be distinguished by a [tʰ] vs. [t] difference.
Even though [tʰ] and [t] phones are phonologically the same for native speakers of English, they may be processed perceptually using different phonetic categories. A recent study (Flege, 1989a; Flege and Wang, 1990) examined the identification of the contrast between English word-final /d/ and /t/ by Chinese subjects. The L1s of these subjects possess a contrast between a voiceless unaspirated /d/ and a voiceless aspirated /t/ in word-initial position, but no Chinese languages has a stop voicing contrast in the final position of words. Both Chinese and native English listeners identified word-final English /t/ and /d/ tokens at high rates. However, when release bursts were removed, only the Chinese listeners showed a significant decrease in performance. This suggested that the Chinese but not the native English subjects were using phonetic categories established for word-initial /t/s and /d/s to identify stops in word-final position. This strategy worked well for word-final stops with bursts but not for those without bursts, which had to be identified on the basis of cues not used to distinguish voiced from voiceless stops in word-initial position e.g., preceding vowel duration.

2.1.3 Phonetic perception

Werker and her colleagues have provided experimental evidence for an optional phonetic level of processing that is distinct from either a phonemic or auditory level (Werker and Logan, 1985; see also Repp, 1981). Other indirect evidence exists for a phonetic level. For example, Koreans who learn English are much more sensitive than native English speakers to the phonetic difference between released and unreleased tokens of word-final /p,t,k/ in English because the presence versus absence of release bursts underlies a phonemic contrast in their L1 (Robson, 1982). However, it is quite easy to make this phonetic difference accessible to English speakers. One has only to direct their attention to it. If listeners did not have access to a phonetic level of representation under certain circumstances it would be difficult to understand how, for example, native speakers of English can detect the subphonemic difference between [tʰ] and [t] realizations of /t/ in standard and Spanish-accented English (Flege and Hammond, 1982). It would also be difficult to explain how adult listeners learn to discriminate certain nonnative phonetic contrasts, such as the contrast between Hindi unvoiced and breathy voiced dental stops (Werker and Tees, 1983), or how listeners may detect socially or stylistically conditioned variations in the realization of an L1 phoneme.
As noted earlier, listeners’ conscious perception of sounds normally occurs at a phonemic rather than a phonetic level. However, it does not appear that phonemic-level processing is obligatory, or takes precedence over phonetic processing in the same way that phonetic-level processing takes precedence over auditory processing of speech (Repp, Healy, and Crowder, 1979). The nature of the processing task is likely to influence whether a phonemic or phonetic mode is engaged. For example, even though consonant perception is often described as “categorical”, within-category discrimination can occur, especially when inter-stimulus intervals are short and the testing paradigm employed minimizes response uncertainty (Carney, Widin, and Viemeister, 1977).

The instructions given to listeners in a speech perception experiment may determine whether they engage an auditory or a phonetic mode of processing (e.g. Repp, 1981; Remez, Rubin, Pisoni, and Carrell, 1981). So too, listeners can be induced to process stimuli in a phonetic rather than phonemic mode. Native speakers of English are accustomed to hearing only voiced and voiceless stop phonemes because English possesses only two phonemic stop series. A study by Pisoni, Aslin, Perey and Hennessey (1982) showed, however, that native English subjects could be trained to partition into three categories the members of a synthetic continuum encompassing stops with lead, short-lag, and long-lag VOT values that would normally be divided into voiced vs. voiceless phonemic categories. The conditions under which access to a phonetic level of representation occurs may vary from listener to listener. They may also vary according to the nature of the listening task, the nature of the stimuli (Logan, Lively, and Pisoni, 1989), and the nature of the phonetic contrast. “Robust” phonetic contrasts that are not found in the listener’s L1 but that occur in many other human languages may be more easily accessible than “fragile” ones (Burnham, 1986).

According to the SLM, phonetic-level perception is mediated by phonetic category “prototypes” which specify the ideal weighting of a set of independent and continuously varying properties that define each phonetic category (e.g. Repp, 1976; Oden and Massaro, 1978; Massaro and Oden, 1980; Massaro and Cohen, 1983, 1984; Samuel, 1977, 1982). Phonetic categories for vowels may specify duration and the patterns of movements to and from “target” formant frequencies. The prototype may be defined as possessing all of the properties associated with a category, each to the largest degree possible. To be identified correctly, an individual token need
not possess all properties, nor must the properties represented be present to the maximum degree possible. For example, a prototypical /i/ may be longer than a prototypical /I/ (Bennett, 1968), but a relatively short [i]-quality vowel may nevertheless be identified as an instance of /i/.

Phonetic category prototypes may evolve so as enable the listener to process the phones of a particular language effectively. By hypothesis, it is with reference to a prototype that listeners determine how a vowel “ought” to sound in a particular context (see Linell, 1982). Also by hypothesis, it is the existence of a tolerance region around each prototype which enables the listener to detect divergences from phonetic norms either as “distortions” (for children or adults with speech defects) or as “foreignness” (in nonnative speakers).

The tolerance region may be greater around vowel than consonant prototypes (Hudgins and Numbers, 1942), perhaps because vowels can vary continuously. However, relatively little work has focussed on defining the range of auditorily acceptable variants for either vowels or consonants.

Phonetic category prototypes seem to evolve as a function of phonetic input even after they have been established and tuned during L1 acquisition (see, e.g., Flege and Eefting, 1986). It appears that the ability to detect foreign accent increases with age (Scovel, 1981). Perhaps listeners become better able to gauge the degree of divergence of particular phones from prototypes as the prototypes defining the “center” of phonetic categories become better defined as the result of experience with an increasingly wide range of variants. Although little work has focussed on second dialect learning, it is my impression that vowels in another dialect of one’s native language sound less accented as a function of experience with the new dialect.

2.2 Perceiving L2 sounds

A large cognitive learning task still awaits the child acquiring L1 phonology after phonetic categories are established and methods are found for implementing them. Children must discern what are the phonemic categories of their L1 (Ferguson, 1986). In keeping with a linguistic approach to speech analysis, Jusczyk (1989) noted that even after children have arrived at a correct description of the phonetic categories in their L1, they still face an additional “mapping problem” that involves relating phonetic categories to the phonological categories of the L1. For native English children this
involves, among other things, learning that [tʰ] and [t] are context-conditioned allophonic variants of /t/ that occur in initial pre-stressed and final post-stressed positions, respectively. Another part of the process of phonological learning is acquiring systematic knowledge of what pattern of sound sequences are permissible in the L1.

Oller and Eilers (1983) found that children just beginning to acquire the phonology of their L1 appear to be better able to discriminate phones that are phonemically contrastive in their L1 than phones which are not. However, the phoneme learning task may be complicated in instances where a phonemic category is implemented by more than one phonetic category. For example, the phoneme /t/ may be implemented as an unaspirated stop or flap in post-stressed intervocalic position. It may be implemented as a lingual stop (usually unreleased) in word-final position, or as a glottal stop. Jusczyk (1985) suggested that variants in complementary distribution (e.g., initial [tʰ] and final [t]) will not be associated in a single phonemic unit until children learn to read.

The learned strategy of focussing attention only on those aspects of sounds needed for phonemic contrasts also seems to characterize the perception of L2 sounds by adults. Most researchers who have examined the production of L2 sounds have recognized the fundamental importance of the L2 learner's attempt to match or find correspondences between phonemic elements in L1 and L2 (e.g., Weinreich, 1953; Briere, 1966; Wode, 1977, 1978; see also Best et al., 1988). Trubetzkoy (1939/1969) hypothesized that the phonology of L1 causes L2 learners to “filter out” acoustic differences that are not phonemically relevant in L1. This view has been restated over the years by many students of L2 learning (e.g., Borden, Gerber, and Millsark, 1983). For example, it is widely agreed that Japanese speakers confuse English /r/ and /l/ perceptually because the phones used to realize these phonemes are not contrastive in Japanese. Logan et al. (1989) observed that learning such a nonnative contrast may be difficult because children’s perceptual sensitivity to speech changes as the result of experience so that only those phonetic contrasts that “denote differences in meaning remain distinctive” (p. 3). This suggests filtering at a phonemic level, whereas the view taken here is that such filtering actually takes place at a phonetic level of processing.

A classic view of perceptual development is that children become increasingly less reliant on sensory information as they develop cognitively (Gibson, 1969). Bruner (1964) suggested three major stages: an enactive
stage, with its reliance on motoric codes; an iconic stage, where reliance shifts to sensory or perceptual codes; and finally a symbolic stage. Bruner (see also Inhelder and Piaget, 1969) suggested that, as children mature, they rely less on the common features which identify specific exemplars as belonging to a category and more on higher order, superordinate, categories in the hierarchy. In keeping with this, the SLM posits that the difficulty adults may have in producing similar L2 sounds authentically is the result of the general cognitive mechanism of equivalence classification.

When asked to imitate an ensemble of familiar and "alien" (i.e., non-L1) synthetic vowels, both children and adults tend to respond with familiar, L1 vowels (Kent and Forner, 1979). Moreover, token-to-token variability in $F_2$ frequency is less for L1 than non-L1 vowels (but cf. Kent, 1974). These imitation data point to the common tendency of listeners to identify vowels in terms of previously established (i.e. L1) categories. Not surprising, researchers have noted that during L2 learning there is a tendency for sounds in the L1 to exert a "gravitational attraction" (Schouten, 1975) or "assimilatory" pull (Best et al., 1988) on L2 sounds. This phenomenon is referred to here as equivalence classification.

Equivalence classification is important for speech learning because it permits listeners to make perceptual groupings of a wide variety of disparate phones with a common communicative function. This basic mechanism, which is evident even in prelinguistic infants (e.g., Kuhl, 1979; Hillenbrand, 1983, 1984), permits humans to perceive constant categories in the face of variability found in the many physical exemplars which may instantiate a category. Without equivalence classification it would be impossible, for example, for talkers to use the word chair correctly in identifying the many physical exemplars of this furniture type.

The role of equivalence classification in speech development seems to find a broad parallel in cognitive development. Children and adults use somewhat different strategies to categorize word, picture, or object arrays (e.g., Anglin, 1977; Bruner et al., 1966; Nelson, 1974). If the development of phonetic category prototypes follows the same general course as concept formation, we would also expect to see an evolution in phonetic perception as children develop. Equivalence classification may underlie the child's ability to produce L1 phones authentically, but it may also be responsible for L2 learners' foreign accent.

The SLM predicts that Late L2 Learners will be less successful in learning similar L2 sounds than Early L2 Learners because they equate similar
L2 Vowel Intelligibility

L2 sounds with sounds in the L1. This may happen even though similar L2 sounds differ acoustically — and audibly in some testing circumstances — from the corresponding L1 sound.

Although it is uncertain why young children should treat similar L2 sounds differently than older L2 learners, one potential basis for the hypothesized difference is that adults and older children may make greater use of higher-order syntactic and semantic information than young children. Auditory-acoustic processing of the speech signal might be terminated prematurely in adults and older children as the result of the relatively rapid recognition of words brought about by greater (or earlier) use of higher order information in parallel with bottom-up phonetic information.

Another possibility is that a difference between young and older L2 learners occurs because of the state of development of phonetic categories at the time L2 learning commences. As children encounter an increasingly wide range of phonetic category realizations, they may become increasingly better able to identify sounds in non-ideal listening conditions because the tolerance limits of the category expand. Although such a development could be regarded as highly adaptive as far as processing of the L1 is concerned, it would make it harder to note phonetic differences between similar L1 and L2 sounds. For example, a native Spanish 5-year-old may be better able to note the acoustic phonetic difference between [t] and [tʰ] phones used to realize /t/ in Spanish and English than a native Spanish adult. At the same time, the center of each category center may become better defined. This may make it easier for children as they grow older to detect distortions and to gauge degree of foreign accent (see Flege, 1988a).

Jusczyk (1989) noted that trying to assign meaning is an important factor in encouraging a child to attend to “similarities and differences that exist in the acoustic attributes” of the speech signal. As children’s lexicons expand to include an ever larger number of minimally paired items, they may perform deeper and more abstract analyses of the acoustic signal, which may influence their readiness to recognize the existence of new phonetic groupings of sounds. For example, 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 were children with relatively poor comprehension abilities.

An increasing tendency to equate similar L2 sounds with sounds in the L1 might be encouraged by greater phonemic awareness, which seems to
increase at about the time children learn to read (Liberman, Shankweiler, Fisher, and Carter, 1974; Bradley and Bryant, 1983). It is still a matter of controversy as to whether an explicit awareness of phonemes arises from learning to read (Morais, Cary, Alegria, and Bertelson, 1979; Mann, 1986) or simply as the result of maturation (Kirtley, Bryant, MacLean, and Bradley, 1989). It is clear, however, that good readers have greater phonemic awareness than poor readers. That is, good readers generally have a greater (and earlier appearing) ability to segment sounds and syllables than poor readers (e.g., Liberman, Shankweiler, Fisher, Liberman, Shankweiler, Fisher, and Carter, 1974; Bradley and Bryant, 1983).

Mann (1984) noted a number of differences between good and poor readers that suggest differences in phonetic processing which, in turn, may be related to a propensity for equivalence classification. She noted that poor readers were less able to remember strings of nonsense words (but not faces or drawings) than good readers. Although both good and poor readers were less able to remember strings of letters that rhymed (e.g., “B”, “D”, “E”) than strings that did not rhyme, the deleterious effect of rhyming was much smaller for poor than good readers in a recall experiment, and also in a study examining the vocal repetitions of sentences consisting of rhyming or non-rhyming words.

Mann (1984) noted that the segmental substitution errors made by poor readers when repeating word strings showed that they, like good readers, code speech phonetically. She was therefore led to conclude that poor readers’ phonetic representations “decay more rapidly” or are “less well formed” (p. 8) than the phonetic representations of good readers. This was supported by a study examining the effect of noise on speech perception (Brady, Shankweiler, and Mann, 1983). High- and low-frequency words were equally intelligible for good and poor readers in ideal listening conditions, whereas poor readers made significantly more identification errors than good readers for words presented in noise. Mann (1984) concluded that good and poor readers differ in terms of how effectively they use phonetic representations to process speech.

The possibility exists that the just-noted reading ability differences between children are analogous to speech processing changes that occur in the ontogenetic development of all individuals; and that differences in phonetic categorization might distinguish individuals, although to a lesser degree, throughout their lifespans. van Balen (1980) found that high school students who received relatively high scores on a test of English listening
proficiency tended also to receive high scores on a comparable test in Dutch. Greenberg and Roscoe (1988) used a list recall task to examine the phonetic coding of speech by students in first-year college foreign-language classes who differed in terms of L2 comprehension ability. A tone or word "suffix" was appended to a list of digits to be recalled. It was assumed that the presence of a word (but not tone) suffix would decrease ability to recall list-final digits more if recall was based more on "vulnerable" sensory information in echoic memory than on "stable, higher-order" phonetic codes. Students who were poor L2 comprehenders showed a significantly greater suffix effect than students with good comprehension ability, suggesting that the good comprehenders were more efficient in their phonetic coding.

Nonnatives comprehend speech more poorly in noise than native speakers (Florentine, 1985). This appears to be due at least in part to the fact that nonnatives' phonetic representations are less well suited for processing incoming L2 phones than native speakers'. Individuals who rapidly code L2 sounds in terms of existing categories, even if they differ substantially from L1 sounds, may succeed well in comprehending the L2, but at the cost of failing to recognize the phonetic differences that may distinguish sounds in L1 and L2.

Paradoxically, individuals who are skilled in comprehending L2 speech in noise, which has been regarded as a suitable index of overall proficiency in the L2 (e.g., Spolsky, Sigurd, Sako, Walker, and Atterbrun, 1968), may experience relatively great difficulty in learning L2 pronunciation because they rapidly and reliably code L2 sounds in terms of existing L1 categories. Lambert (1977) suggested that some L2 Learners ("code users") are likely to perceive an L2 sound which differs auditorily from sounds in the L1 in terms of L1 categories, whereas others ("code formers") tend to develop new central representations in such instances. Such differences might ultimately be used to account for why some adults profit more from auditory perceptual training on a novel phonetic contrast than others (e.g., Flege, 1989a; Flege and Wang, 1990).

The SLM differs from previous approaches in not regarding equivalence classification as performing a kind of affective, auditory, or phonological "filtering" of subphonemic acoustic differences between L1 and L2 sounds. By hypothesis, all audible acoustic differences between similar L1 and L2 sounds may influence the phonetic system, even those that are not accessible perceptually (that is, those that cannot be heard consciously because of previous perceptual learning). This can be illustrated by consid-
erating English voiceless stops from the standpoint of native speakers of French. Flege (1984) found that native French adults who were highly experienced in English produced English /t/ with significantly longer VOT values in English than in French, but with significantly shorter VOT values than native speakers of English. By hypothesis, they were prevented from producing English /t/ authentically because they had not established a long-lag stop category for English /t/. Their phonetic category prototype for /t/ had probably changed, however, because of the many English [tʰ] phones they had identified as /t/.

The hypothesis that already-established (L1) categories may evolve as a function of L2 experience in the direction of L2 sounds agrees with the observation by Obler and Albert (1978: 159) that bilinguals identify phonemes “without regard for language specific (acoustic) information” because their perceptual processing is “intermediate to that of...monolinguals”. This may not be true for all L1 categories and all L2 sounds, however. An hypothesis of the SLM is that L1 categories will not be modified in an attempt to accommodate the learners need for producing and perceiving L2 sounds referred to as “new”.

2.3 The new vs. similar distinction

The distinction between new and similar L2 sounds has been made before. 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 phonic 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 a crucial (but undefined) range are judged to be non-equivalent, and will undergo “other developments” than simple substitution, according to Wode.

The SLM posits that the basis for a sensitive period is the increasing tendency by older children and adults to classify as similar sounds in the L2 that would be classified as new by young children. One might characterize L1 learning by young children as a “bottom-up” process of learning, whereas L2 learning by older children and adults might better be charac-
terized as a "top-down" process (Mack, 1989). For children learning their L1, all sounds are new. The number of phonetic categories they will establish depends on the number of sounds (i.e. phone classes) encountered in the L1. For example, English-learning children will establish many more vowel categories than Spanish-learning children. Older children and adults who are learning an L2, on the other hand, have already established a phonetic system suitable for distinguishing a large and ever growing number of lexical items. The number of additional categories they establish will be limited by their previous phonetic learning via the mechanism of equivalence classification.

2.4 Phonetic norm

Many studies of L1 speech learning, both those comparing groups in cross-sectional designs (e.g. Smith, 1977) or examining single subjects in detail in longitudinal designs (Mack and Lieberman, 1985), have assumed that children's speech evolves towards something referred to as a "phonetic norm". The L1 norm, which should probably be regarded as a heuristic for research rather than a reality, is typically defined as monolingual adults' mean values for various phonetic parameters of interest. The adults chosen to represent the norm that children are said to be approximating are typically drawn from the same community as the children being studied.

The concept of phonetic norm is also used frequently in studies of L2 speech learning as a benchmark for assessing the extent to which L2 learners have succeeded in producing L2 sounds, and as a starting point for assessing the phonetic distance between sounds in L1 and L2. Defining the L2 phonetic norm can be problematical. For example, Holden and Nearey (1986) raised the issue of how one is to chose among the "widely divergent dialects" of an L2 in defining the L2 phonetic norm.

If one is interested primarily in determining how successful learning has been, the L2 norm should probably be based on all L2 speech that has been processed meaningfully. For many L2 learners, this would include the speech of other nonnative speakers who speak the L2 with a foreign accent, and native speakers who are representative of various regional and social dialects in addition to those who use the standard or prestige dialect. The norm should also reflect the variations in L2 speech typical of varying levels of formality.

For a variety of practical and methodological reasons, the L2 norm has seldom if ever been defined in an ideal fashion. An additional methodolog-
ical problem is the need to relate an individual L2 Learner’s performance to that of a group of speakers of the L2 (i.e., the L2 norm). For example, a Spanish speaker whose typical realization of /i/ is an [I]-like vowel might have greater difficulty in learning English /I/ than a Spanish speaker who typically realizes /i/ as an [i]-quality phone. Holden and Nearey (1986) showed that which Russian vowel was identified with English /ʌ/ varied according to the dialect of the native Russian listener.

A particular listener’s prototype for a vowel might be estimated by presenting a wide range of synthetic variants differing according to the acoustic features that distinguish adjacent vowel categories (see Holden and Nearey, 1986). The phonetic norm for a language could then be based on data obtained for representative native speakers of that language. However, this would be an enormous undertaking, even for a single subject, because of the number of combinatorial possibilities for the relevant acoustic parameters in vowels (e.g., duration, $F_0$, $F_1$, $F_3$ “target” frequencies and movement patterns).

Speech production and perception are related to one another, at least for individuals with stable phonetic systems. Because of this, the prototypes of a language might be estimated indirectly, through an examination of speech production. A simple (but not necessarily optimal) way of doing so is by plotting target formant frequency values at a single measurement point for individual tokens (or means for individual talkers) in an $F_1$-$F_2$ space (see e.g., Holtse, 1972). Motivation for this is provided by multidimensional scaling experiments which have shown that $F_1$ and $F_2$ are independent dimensions that account for most of the variance in spectral quality ratings for vowels in a relatively large vowel inventory (Pols, van der Kamp, and Plomp, 1969). The perceptual prototype can be estimated as the value found at the center of the area thus delimited.

When plotted in this way, the realizations of vowels adjacent to one another in the phonetic space often overlap (Peterson and Barney, 1952) suggesting that vowels may be confused perceptually if represented only by $F_1$ and $F_2$ (Carlson, Granström, and Fant, 1970; DiBenedetto, 1989). The majority of vowel tokens can nevertheless be classified correctly on the basis of the $F_1$ and $F_2$ frequency values by statistical procedures. The rate of correct classification can be increased by normalization, and increased further still by taking into account vowel duration and the pattern of dynamic formant movements in the first two formants (Assman et al, 1982; DiBenedetto, 1989; see also Miller, 1989 for three-dimensional plots).
For the purposes of L2 research, a perceptual "tolerance region" for vowels in the L2 can be defined as some portion of the space defined by the vowels produced by native speakers, for example, a space encompassing a 95% confidence interval along the principal components of variation of F₁ and F₂. Other, correctly identifiable, tokens produced by nonnative talkers that are foreign-accented may fall outside the tolerance region defining a vowel category.

Practical considerations dictate that only a small number of native speakers can be examined to estimate the phonetic norms of the L1 and a target L2 being learned. This raises the important issue of deciding which native speakers (and which speaking style) should represent the phonetic norm. In most previous studies, the talkers chosen to represent L2 have been drawn from the community in which the L2 learners reside. The L1 norm has often been based on L2 learners' productions of their own native language, a practice that may bias results if learning the L2 has affected L1 production. The style examined in most L2 research is usually the maximally careful style elicited in formal production experiments, which may limit the generalizability of findings.

2.5 The relationship of L2 and L1 sounds

A useful method for characterizing the relationship between phonetic categories in L1 and L2 is to classify the L2 sounds as "new", "similar", or "identical". A three-way classification of L2 sounds is implicit or explicit in much L2 research (e.g., Brière, 1966). No universally-accepted method now exists for differentially classifying L2 sounds as new or similar. In attempting to operationalize the distinction, the most important question to consider is: When does an L1 versus L2 acoustic difference make a phonetically relevant difference?

To determine this, the SLM employs three criteria for classification. A preliminary step is to consider the IPA symbols used to represent the L1 and L2 sounds. This is followed by acoustic measurements and listeners' perceptual judgments of sounds in L1 and L2. The SLM posits that interlingual identification occurs at a phonetic rather than a phonemic level, so the procedures operate on sounds (that is, phonetically-relevant phone classes).

An identical L2 sound is a sound represented by the same IPA symbols used to represent a sound in the L1. When acoustic analyses are performed
for representative native speakers, there is not a significant acoustic difference between corresponding L1 and L2 sounds. When a detailed perceptual analysis is performed, listeners cannot detect a difference between the L1 and L2 sounds. An identical L2 sound is usually produced authentically as the result of a process referred to as "positive transfer" (Weinreich, 1953). Identical sounds have therefore received little attention because most L2 speech errors involve similar and new sounds (James, 1984, 1985a). However, for consonants at least, positive transfer may not extend to new syllable positions. Recent research has shown that Chinese subjects have difficulty producing and perceiving a contrast between /t/ and /d/ in the final position of English words even though an English-like /t/-/d/ contrast occurs in the initial position of words in their L1 (Flege and Davidian, 1985; Flege et al, 1987; Flege, 1988c).

To be classified as either similar or new, some acoustic difference(s) between pairs of L1 and L2 sounds must exist, and there must be evidence that the sounds are auditorily discriminable. A "phonetic symbol" criterion might be used as a provisional measure because, at present, no well-accepted, objective metric exists for measuring the phonetic distance between sounds in two language. An L2 sound that is similar to a sound in L1 is represented by the same IPA symbol as the L1 sound, even though statistical analyses reveal significant — and audible — acoustic differences between the two.

For example, Flege (1987a) examined formant frequency values in French /i/ and English /i/, and also between the /u/s of French and English. Significant acoustic differences were obtained for both pairs of French-English vowels; and native English listeners were able to identify French vowels at significantly above-chance rates when asked to determine which member of a vowel pair had been spoken by a native French speaker. Both the acoustic and perceptual tests conformed to traditional phonetic descriptions of French /i/ and /u/ being more "peripheral" than their English counterparts.

An L2 sound that is new differs acoustically and perceptually from the sound(s) in L1 that most closely resembles it. But, unlike a similar sound, it is represented by an IPA symbol that is not used for any L1 sound. An example of a new sound from the standpoint of English is French /y/. This vowel sound differs acoustically and perceptually from the nearest possible vowels of English (/i/, /I/, and /u/), and is represented by a symbol not used traditionally in describing the vowels of English.
The phonetic symbol criterion has the advantage of making use of the expert phonetic classifications of researchers who have worked with the languages under consideration, but it is not without problems. The most obvious problem is that many phonetic transcription systems are now in use; and even seasoned researchers who are using the nominally same system don’t always agree. For example, the distinction between the vowels in English words like *beat* and *bit* is sometimes represented as a distinction between /i/ and /I/, and sometimes as one between /i:/ and /i/. The latter symbolization, which emphasizes the duration difference that accompanies the spectral distinction between this tense-lax vowel pair, seems to be favored by analysts whose L1 makes important use of duration for phonemic distinctions.\(^7\)

Thus it appears necessary to supplement the phonetic symbol test with additional acoustical criteria. Bohn and Flege (1991) suggested that an L2 vowel should be considered new only if most of its realizations occupy a portion of the acoustic phonetic vowel space that is unoccupied by the realizations of any L1 vowel. This implies that few of the vowels in an L2 will be new for learners whose L1 has a large vowel inventory. Perceptual tests might also be used to differentiate new and similar vowels.\(^8\) Other behavioral measures of speech processing might also furnish a useful metric.\(^9\)

2.6 Applying the classification scheme

Application of the criteria is illustrated in Figure 1(a-d) for four hypothetical languages. In these illustrations, the simplifying assumption is made that the languages possess only the small number of vowels shown, and that non-spectral dimensions such as duration, diphthongization or nasalization are not used contrastively.

Figure 1(a) represents an L1 with three vowel phonemes (/i/, /u/, /a/) and an L2 with four (/i/, /u/, /a/, /e/). The vowel /i/ occupies approximately the same portion of the acoustic vowel space in L1 and L2; the L2 /i/ is classified as identical because the small and nonsignificant acoustic differences between it and the L1 /i/ are inaudible. The L2 /a/ is classified as identical for the same reasons. The L2 /u/, on the other hand, is classified as similar because it is realized as a slightly lower [u]-quality vowel (viz. [u]) than the L1 /u/, even though its major (and only) allophone is represented by the same symbol used for the major allophone of L1 /u/ (viz., /u/). The L2 /e/
is classified as new because, in addition to differing acoustically and auditorily from adjacent L1 vowels, it is represented by a symbol not used for a major allophone for any of the L1 vowels.

Figure 1(b) illustrates how allophonic variation is to be handled. Once again, the L1 has three vowel phonemes (/i/, /u/, /a/) and the L2 has four (/i/, /u/, /a/, /ε/). However, in this instance one of the L1 vowel phonemes, /a/, has two allophonic variants ([ε], [a]) that are acoustically and auditorily indistinguishable from the major allophones used to implement the /ε/ and /a/ phonemes of L2. Since interlingual identification is hypothesized to occur at a phonetic category level, the L2 /ε/ and /a/ are classified as identical vowels.

As noted earlier, conscious speech perception normally occurs at a phonemic level, so the identity of the L1 and L2 vowels might not be apparent immediately. If so, a learner might produce an L2 word like /ba/ as [bε] in early stages of learning because /a/ is realized as [ε] in open syllables. The SLM predicts, however, that this phonemically-motivated substitution pattern will not persist because interlingual identification occurs at a phonetic category level. This leads to the expectation that learners will eventually use their [ε] and [a] phonetic categories to implement the L2 /ε/ and /a/ phonemes.

The allophonic relationship between the [ε] and [a] phones in the example just given can be regarded as learned "phonological" knowledge. Steensland (1981) found that, when Swedes were asked to identify Russian /e/ in terms of Swedish vowel categories, the Russian /ε/ more closely approximated Swedish /ε/ than Swedish /e/ when it was produced in the context of palatal than plain consonants in Russian. These phonetically-conditioned differences in vowel height would presumably have been less obvious to native speakers of Russian had some means been found to test their perception. For example, Jaeger (1986) used the concept formation paradigm to show that native speakers of English group [kh] and [k] phones into a single /k/ phonemic category despite the fact that they differ audibly.

The claim that such phonological knowledge will not affect L2 production after an initial stage of learning may need to be restricted to vowels; Benson (1988) found that Vietnamese subjects had no difficulty producing consonants such as /p,t,k/ in the final position of English words except those containing the diphthong /a/. In this context final stops are not permitted in Vietnamese. The fact that the subjects honored L1 phonotactic constraints in English L2 suggests, indirectly, that they perceived the vowel in English
Figure 1(a-d). Illustration of how vowels in four hypothetical L1-L2 pairs are classified using procedures outlined in the text.

**Figure 1a**  
**F2 Frequency in Mels**

<table>
<thead>
<tr>
<th>L1 phoneme realized as</th>
<th>context</th>
<th>L2 phoneme realized as</th>
<th>context</th>
<th>L2 sound classified as</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>[i]</td>
<td>all</td>
<td>/i/</td>
<td>[i]</td>
</tr>
<tr>
<td>/u/</td>
<td>[u]</td>
<td>all</td>
<td>/u/</td>
<td>[u]</td>
</tr>
<tr>
<td>/a/</td>
<td>[a]</td>
<td>all</td>
<td>/a/</td>
<td>[e]</td>
</tr>
</tbody>
</table>

**Figure 1b**  
**F2 Frequency in Mels**

<table>
<thead>
<tr>
<th>L1 phoneme realized as</th>
<th>context</th>
<th>L2 phoneme realized as</th>
<th>context</th>
<th>L2 sound classified as</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>[i]</td>
<td>all</td>
<td>/i/</td>
<td>[i]</td>
</tr>
<tr>
<td>/u/</td>
<td>[u]</td>
<td>all</td>
<td>/u/</td>
<td>[u]</td>
</tr>
<tr>
<td>/a/</td>
<td>[a]</td>
<td>all</td>
<td>/a/</td>
<td>[a]</td>
</tr>
</tbody>
</table>
Figure 1c  \( F_2 \) Frequency in Mels

<table>
<thead>
<tr>
<th>Phoneme Realized as</th>
<th>Context</th>
<th>Phoneme Realized as</th>
<th>Context</th>
<th>L₂ Sound Classified as</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>[i]</td>
<td>all</td>
<td>/i/</td>
<td>[i]</td>
</tr>
<tr>
<td>/u/</td>
<td>[u]</td>
<td>all</td>
<td>/u/</td>
<td>[u₂]</td>
</tr>
<tr>
<td>/ɛ/</td>
<td>[ɛ]</td>
<td>all</td>
<td>/ɛ/</td>
<td>[ɛ]</td>
</tr>
<tr>
<td>/æ/</td>
<td>[æ]</td>
<td>all</td>
<td>/a/</td>
<td>[a]</td>
</tr>
</tbody>
</table>

Figure 1d  \( F_2 \) Frequency in Mels

<table>
<thead>
<tr>
<th>Phoneme Realized as</th>
<th>Context</th>
<th>Phoneme Realized as</th>
<th>Context</th>
<th>L₂ Sound Classified as</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>[i]</td>
<td>all</td>
<td>/i/</td>
<td>[i]</td>
</tr>
<tr>
<td>/u/</td>
<td>[u]</td>
<td>all</td>
<td>/u/</td>
<td>[u₂]</td>
</tr>
<tr>
<td>/ɛ/</td>
<td>[ɛ₁]</td>
<td>all</td>
<td>/ɛ/</td>
<td>[ɛ]</td>
</tr>
<tr>
<td>/æ/</td>
<td>[æ]</td>
<td>all</td>
<td>/a/</td>
<td>[a]</td>
</tr>
</tbody>
</table>
words such as *bite* as [a³], not as a related Vietnamese vowel that might have occurred in a consonant-final word (e.g. /ʌ/).

Figure 1(c) represents a kind of phonemic merger, that is, an instance in which the L1 has a phonemic contrast not found in the L2. The L2 phoneme is represented here as /EA/ because the wide range of phones used to realize it have (hypothetically) lead researchers to transcribe it as either /ɛ/ and /æ/. The simple substitution of an L1 /ɛ/ for the L2 /EA/ will sometimes result in correct identification by a native L2 listener, but sometimes it will not. Whether or not correct identification occurs will depend upon the nature of L2 listeners' prototypes for /EA/. So too, the simple substitution of an L1 /æ/ for /EA/ will sometimes but not always result in correct identification. The L2 /EA/ is classified as similar, but it is not obvious here whether it is similar to L1 /ɛ/ or to L1 /æ/. A post-hoc analysis of L2 production data could reveal this: A learner who substitutes L1 /ɛ/ is likely to have identified the L2 /EA/ with L1 /ɛ/, whereas the learner who substitutes L1 /æ/ probably identified /EA/ with L1 /æ/.

Finally, Figure 1(d) illustrates a kind of phonemic split, that is, an instance in which the L2 has a phonemic contrast not found in the L1. It underscores the point that the classification of L2 sounds must be made with reference to the individual speaker-hearer. The /ɛ/ and /æ/ phonemes of L2 are realized with phones that might be used to realize a single phoneme in L1, represented here as /EA/. The hypothetical L1 /EA/ is realized by a wide range of variants, [æ]-quality vowels by older talkers and [ɛ]-quality vowels by young talkers. A young native speaker of L1 who realizes /EA/ as [ɛ] and uses the L1 [ɛ] category to produce L2 /ɛ/ should be understood. For such an individual, the L2 /ɛ/ can be classified as an identical vowel if the learner's L1 [ɛ] realizations are judged to be "good" exemplars of L2 /ɛ/. The L2 /æ/ would necessarily be regarded as a new vowel. Just the opposite set of L1-L2 relationships would hold true for a hypothetical older talker who realized the L1 /EA/ as an [æ]-quality vowel.

3. Predictions of the model

Early L2 learners are tentatively defined as individuals who begin learning the L2 by the age of 5-6 years; Late L2 learners are defined as those who begin learning their L2 later in life. The SLM posits that phonetic categories are needed for the authentic production of L2 sounds. By
hypothesis, Early L2 Learners are able to establish phonetic categories for all L2 sounds not found in the L1, and so possess an enriched phonetic system which includes the phonetic categories possessed by monolingual speakers of both the L1 and the L2. This leads to the hypothesis that:

H3: Early L2 learners will produce identical, similar, and new L2 sounds authentically if they have received sufficient phonetic input.

The SLM posits that Late L2 learners will differ from Early L2 Learners with respect to similar sounds but not identical or new sounds. This position diverges from the one taken by some (but not all) previous investigators. Stockwell and Bowen (1965) felt that the absence of an L2 sound in the L1 was the source of the greatest learning difficulty (see also Olson and Samuels, 1973), whereas the partial resemblance of an L2 sound to a sound in the L1 generally facilitates learning. Based on the results of a short-term training experiment, Brière (1966: 795) concluded that L2 sounds which are "close equivalents" of L1 sounds at either a phonemic or a phonetic (i.e., "allophonic") level will be "easier to learn" than L2 sounds without such equivalents. The position taken by Oller and Ziahosseiny (1970), on the other hand, was that L2 sounds that differ only minimally from sounds in the L1 are the most difficult to master.

There may be an element of truth to both positions. Snow and Hoenen-Höhle (1982) report that those Dutch sounds that showed relatively little improvement over 20 trials in a laboratory study evaluating imitation by native English adults and children showed a relatively great amount of improvement during a one-year period of naturalistic acquisition of Dutch in the Netherlands. L2 sounds with close equivalents in the L1 may be relatively easy to produce in the earliest stages of L2 learning whereas L2 sounds which are more dissimilar are relatively difficult because they require new gestures to be produced in an acceptable — or even an identifiable — fashion. Relatively greater success for similar than new sounds is what the SLM predicts for students in classroom learning situations where native speaker input is unavailable or limited; it might also be seen in early stages of naturalistic acquisition for learners who have native-speaker input. However, the SLM proposes that the reverse is true for later stages of naturalistic L2 learning.

The SLM posits that Late L2 Learners can establish phonetic categories for new L2 sounds whereas they can not do so for similar L2
sounds. Thus the SLM leads to a prediction that L2 learners will ultimately be more successful in producing L2 sounds authentically if they differ substantially from L1 sounds than if they differ just a little.

By hypothesis, Late L2 learners may at first identify new sounds with some sound(s) in the L1, but this will not persist because the new L2 sound will not be equated with an L1 sound. It is predicted that Late L2 learners will, given sufficient phonetic input, note the phonetic difference between the new L2 sound and any sound in L1, and that this will precipitate the establishment of a phonetic category for the new L2 sound. Once a realization rule has been developed with which to output the newly established L2 category, production of the L2 sound will be authentic.

In regards to the experiment to be presented below, experienced speakers of Dutch were expected to produce new English vowels more authentically than inexperienced ones. Equivalence classification is hypothesized to be the primary obstacle to effective speech learning beyond early childhood for similar sounds, but no such barrier should exist for new vowels. Thus the intelligibility of new vowels produced by experienced Dutch subjects should be as great as that for native speakers of English. This can be formalized as:

H4: In early stages of learning, Late L2 learners may fail to produce a new L2 sound authentically, substituting one or more L1 sounds for it. However, experienced Late L2 learners will eventually develop phonetic categories for new L2 sounds and produce them authentically.

The SLM leads to a radically different expectation concerning Late L2 learners’ production of similar L2 sounds. The SLM posits that Late L2 Learners continue to equate similar L2 sounds with a sound(s) in the L1 no matter how long they have spoken the L2 and now matter how good they are at learning languages.

By hypothesis, equivalence classification places an upper limit on the extent to which L2 learners can approximate the phonetic norms of L2 for similar vowels (Flege, 1981, 1984, 1988b; Flege and Hillenbrand, 1984) by preventing them from establishing a phonetic category for the similar L2 sound. So, for example, French ten-year-olds but not five-year-olds are expected to equate English [tʰ] and French [t] realizations of /t/. If so, then ten-year-old learners should at best phonetically approximate the VOT norm for English /t/ even though they are able to detect auditorily the
acoustic difference between English and French /t/ realizations. They are expected to realize /t/ with longer VOT values in English than French, but this would result from the use of different realization rules, not from implementing the phoneme /t/ using two different phonetic categories.\(^\text{10}\) This leads to the formalization of two additional hypotheses:

**H5:** Late L2 learners will not produce similar L2 sounds authentically, no matter how experienced they are, because category formation is blocked by equivalence classification.

**H6:** Experienced Late L2 learners will differ from inexperienced Late L2 learners to a greater extent for new L2 sounds (for which categories can be established) than for similar L2 sounds.

These hypotheses generate the prediction that Dutch subjects, even those with an excellent overall pronunciation of English, will continue to show phonetic interference in producing similar L2 sounds.

Determining the exact age at which L2 Learners are no longer likely to establish an additional phonetic categories for similar L2 sounds (i.e., the age of demarcation between Early vs. Late L2 Learners) will require additional research. It should be noted that the age proposed here, 5-6 years of age, differs greatly from the age proposed by Best et al., (1988). The authors suggested that by the age of 10-12 mos, infants begin categorizing sounds phonemically. As a result, they may lose the ability to discriminate two novel but phonetically distinct sounds classified as allophonic variants of an L1 phoneme. Two phonetically distinct sounds which are not “assimilated” to an L1 phoneme will remain discriminable even though neither occurs in the L1. This was the explanation given for why children being reared in an English-speaking environment, and English-speaking adults, remained able beyond the age of 10-12 months to discriminate Zulu clicks. According to Best et al., such non-assimilating (i.e., new) sounds may continue to be perceived in an auditory or phonetic (articulatory) mode.

4. Vowel production data

Relatively few previous studies of L2 learning have provided data suitable for evaluating the SLM’s predictions concerning vowel production. An /e/-for/-æ/ substitution reported for Bulgarian-accented English by Danchev (1987) at first seems to contradict the SLM’s predictions. Bulgarian is
analyzed as having /e/ and /æ/ phonemes but no /ɛ/ (Maddieson, 1984), leading one to think that /æ/ is a new vowel and will therefore be produced authentically by Bulgarians who are experienced in English. Danchev (1986) reported, however, that Bulgarians substitute their /e/ (and, less frequently /ɑ/) for English /æ/. He noted that an /e/-for-/æ/ substitution may be “consciously or unconsciously accepted” by native English listeners and Bulgarian teachers of English because of the acoustic overlap seen between /ɛ/ and /æ/ in British English (henceforth, BE).

The substitutions in Bulgarian-accented English probably do not represent a disproof of the prediction about new vowels, however. First, Danchev’s (1986) observations were based on several sources of data which included spelling errors and loanword phonology as well as actual pronunciations by L2 learners. The L2 learners were high school and university students and individuals enrolled in intensive English classes who may not have fulfilled the condition of “sufficient native speaker input”. English-learning children must receive several years of input before they succeed in producing /æ/ (Amastae, 1978; Mack and Lieberman, 1985; Pollock and Keiser, 1990), so there is no reason to expect Bulgarian adults to do so on the basis of less input.

The predictions regarding L2 vowel production generated by the SLM received some support from two instrumental studies of vowel production. One (Flege, 1987a) measured formant frequencies to test the difference between similar and new vowels in French, viz. /u/ and /y/. Previous research had suggested that English speakers substitute English /u/ for the new French vowel /y/ in early stages of L2 learning, but little was known concerning how experienced English speakers of French would perform. The most experienced of the three native English groups examined did not differ from French monolinguals in producing the new vowel /y/, whereas all three groups differed from French monolinguals in producing the similar vowel /u/.

Perceptually-based results obtained by Major (1987) suggested that new vowels may be learned more successfully than similar vowels. That study examined the production of English /æ/ and /ɛ/ by 50 Brazilian subjects. From the standpoint of Portuguese, English /æ/ is apparently a new vowel and English /ɛ/ is a similar vowel. Accordingly, the SLM predicts that Portuguese learners of English will establish a phonetic category for English /æ/, and eventually produce that vowel authentically, whereas they will be unable to establish an /ɛ/ category and thus will continue to produce that vowel nonauthentically.
Native speakers of English in the Major (1987) identified /æ/ (in sat) and /e/ (in bet) in a two-alternative forced-choice test. The percentage of correct identifications was calculated for two groups of Brazilian subjects differing in overall degree of foreign accent. The relatively non-proficient subjects’ /e/ was more intelligible than their /æ/ (87% vs. 26% correct identifications), whereas the reverse was true for the proficient subjects (50% vs. 68% correct). These differences led to a significant Group x Vowel interaction in an ANOVA examining the percent correct scores ($p < .01$). It appeared that as the Brazilians were learning English /æ/ their production of English /e/ somehow deteriorated, perhaps because they over-generalized their solution for a known pronunciation problem. A definite conclusion cannot be reached, however, because the listeners were offered only two choices, which means that results for one vowel necessarily affected results for the other vowel.

The study presented below also used a perceptual method to assess the authenticity of L2 vowel production, but it made use of 14 rather than just two response choices. The number of response categories is just one of many factors that may influence how well vowels are identified in an identification experiment. Some of these factors will be discussed before the English vowels examined in the present study are classified (as identical, similar, or new) and predictions generated concerning Dutch subjects’ ability to produce them.

5. Factors affecting vowel intelligibility

Intelligibility is determined by how a talker articulates and also by the interlocuter’s listening skill. For example, the English spoken by deaf talkers is more intelligible for listeners familiar with deaf individuals, such as teachers of the deaf, than it is for listeners who are not familiar with deaf speech (McGarr, 1978). The effect of listener experience appears to be greater for isolated words than sentences, and greater in testing paradigms with high than low uncertainty (see Rubin, 1983). One study (Rubin, 1983) did not show the expected advantage of previous deaf listening experience, but this was probably because the deaf talkers were presented in a variable-talker rather than a constant-talker format (see below).

Generally speaking, the greater the use of higher order information in reaching phonetic decisions, the more intelligible vowels will be (Miller,
Heise, and Lichten, 1951). Kalikow, Stevens, and Elliot (1977) showed that words presented in noise were identified better when they occurred in a semantically constraining context than at the end of sentences like *John was discussing the__*. Varonis and Gass (1982) had native English listeners rate sentences produced by non-native speakers. Both grammatical errors and non-authentic pronunciation affected judgments of comprehensibility. Gass and Varonis (1984) found that familiarity with the topic of conversation had an important effect on listeners' comprehension; familiarity with particular talkers and foreign accents were also important.

Not all of the English vowels spoken by native English talkers are identified correctly by native English listeners, so one would not expect perfect intelligibility for nonnative talkers, even if they produced English vowel in a completely native-like way. Table 1 summarizes the results of several previous studies that have assessed the intelligibility of the six English vowels examined in the present study. The overall rate of correct identifications was greater for high vowels (/i/, /ɪ/, /u/), which were identified at near-perfect rates, than for non-high vowels (/æ/, /a/, /ʌ/), which were not. The misidentifications of the non-high vowels occurred even in studies employing

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/ (100)</td>
<td>/i/ (99+)</td>
<td>/i/ (99)</td>
<td>/u/ (100)</td>
<td>/i/ (99+)</td>
<td>/i/ (100)</td>
<td>/i/ (98.8)</td>
<td></td>
</tr>
<tr>
<td>/i/ (100)</td>
<td>/a/ (99+)</td>
<td>/i/ (99)</td>
<td>/i/ (99)</td>
<td>/u/ (99+)</td>
<td>/i/ (97)</td>
<td>/i/ (98.7)</td>
<td></td>
</tr>
<tr>
<td>/u/ (100)</td>
<td>/i/ (99)</td>
<td>/a/ (96)</td>
<td>/a/ (99)</td>
<td>/æ/ (98)</td>
<td>/a/ (97)</td>
<td>/a/ (98.2)</td>
<td></td>
</tr>
<tr>
<td>/æ/ (100)</td>
<td>/a/ (97)</td>
<td>/a/ (96)</td>
<td>/i/ (98)</td>
<td>/i/ (96)</td>
<td>/a/ (96)</td>
<td>/a/ (95.8)</td>
<td></td>
</tr>
<tr>
<td>/ɪ/ (93)</td>
<td>/ɪ/ (96)</td>
<td>/ɪ/ (96)</td>
<td>/æ/ (97)</td>
<td>/a/ (96)</td>
<td>/æ/ (90)</td>
<td>/æ/ (95.2)</td>
<td></td>
</tr>
<tr>
<td>/a/ (89)</td>
<td>/æ/ (94)</td>
<td>/æ/ (92)</td>
<td>/ʌ/ (87)</td>
<td>/ʌ/ (92)</td>
<td>/ʌ/ (78)</td>
<td>/ʌ/ (90.3)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Percent correct identification by native English listeners of vowels spoken in CVC words by American and Canadian talkers. The vowels in each of six previous studies (A-F) were identified in fixed-talker conditions using a variable number of response categories (keywords).

A. Macchi (1980): 11 vowels in a /t_t/ context, 15 talkers, 96 listeners
B. Strange (1989): 11 vowels in /b_b/, /d_d/ and /d_d_t/ contexts, 1 talker, 36 listeners
C. Strange (1989): 10 vowels in six CVC contexts, one talker, 112 listeners
D. Assman et al. (1982): 10 vowels in a /p_p/ context, 10 talkers, 14 listeners
E. Strange et al. (1976): 9 vowels in a /p_p/ context, 15 talkers, 11 listeners
F. Strange and Gottfried (1980): 10 vowels in a /k_k/ context, 6 talkers, 20 listeners
listeners who had been carefully screened and/or trained. When misidentifications occurred, /æ/ was typically heard as /e/ or /ʌ/, /a/ was heard as /ʌ/, and /ə/ was heard as /a/ or /æ/.

The misidentifications of /æ/, /a/, and /ʌ/ probably derived from the relative lack of peripheral auditory discriminability. Sinnott (1989) found that although human subjects discriminated perfectly pairs drawn from a synthetic set of 10 English vowels based on Peterson and Barney (1952) values, they were relatively slow in doing so for /a/-/ʌ/, /æ/-/e/, and /ɛ/-/ɪ/. Non-human primates were slow for the same pairs. In addition, the animals showed errors in discriminating /a/-/ʌ/, /ɛ/-/ɪ/, and /æ/-/e/. It is notable that the macaques made far more discrimination errors for /a/-/ʌ/ (20.5%) and /æ/-/e/ (12.0%) than for /ɪ/-/ɪ/ (0.5%).

The importance of vowel identity on intelligibility can be illustrated by considering the results obtained by Eguchi and Hirsh (1969). The /i/, /ɛ/, and /æ/ stimuli in that study had been edited from the sentence He has a blue pen as spoken by 84 adults and children. As shown in Figure 2, the correct identification rates were far from perfect for any group, but they were higher for adults and older children than for children aged 3-5 years. The relatively low intelligibility of the young children’s vowels was probably due to their vowel articulation rather than to the (adult) listeners’ inability to normalize for formant frequency differences between adult- and child-produced vowels, for token-to-token variability in F1 and F2 frequencies decreased steadily with age.

The Eguchi and Hirsch (1969) data in Figure 2 show that /i/ and /ɛ/ were identified more often than /æ/ (just 9% correct), which was often heard as /e/ because of allophonic, style-conditioned variation. Increases in speaking rate are well known to decrease the rate of information transmission in many communication systems, including fingerspelling (Reed, Delhorn, and Durlach, 1987). The temporal and spectral contrasts between vowels are reduced as speech rate is increased (Koopmans van Beinum, 1980), leading to decreased intelligibility. It appears that a relatively slow speaking rate may facilitate the comprehension of an L2 by nonnatives (Koster, 1987).

Most studies of L2 production have examined isolated words that have been read from a list. This virtually assures that fast-rate effects will not lessen intelligibility. However, speaking rate may influence the intelligibility of English vowels spoken by some nonnative speakers in real-life situations. Anderson-Hsieh and Koehler (1989) found that English passages produced
Figure 2. The intelligibility of the vowels /i/, /e/, and /æ/ edited from an English sentence spoken by child and adult native speakers of English. One group of listeners (top) used six response categories and another group of listeners (bottom) used 12 response categories. The data are from Eguchi and Hirsh (1969).
at a fast rate by native Chinese subjects with poor English pronunciation were much less comprehensible than their normal-rate sentences. The fast-rate decrement noted for a subject with good English pronunciation was much smaller, resembling that of a native English speaker.

Including the phonetic context that originally surrounded a vowel usually makes it easier to identify. For example, Rubin (1983) found that vowels in /bVb/ words were identified 3% more often when the words were presented at the end of their carrier phrase than when removed from the carrier phrase, and about 7% more often in the original /b_b/ context than when presented in isolation. Phonetic context may also aid the listener in identifying vowels spoken by nonnative speakers. Amastae (1978) transcribed English vowels read by nine native speakers of Mexican Spanish. More vowels from a word list were produced correctly than vowels read in a text (95% vs. 89%).

Consonantal context has been described as a factor which “perturbs” the acoustic structure of vowels (Stevens and House, 1963). This implies that vowels might have invariant acoustic properties if it were not for coarticulation; and that coarticulation impedes identification. More recent studies examining vowels in CVC syllables have shown that consonant context may actually aid vowel recognition. When presented in isolation, the “target” portion of vowels can be identified at high rates (Assman et al, 1982). However, vowels may also be identified quite well on the basis of the transitions leading to and from the target (see Rubin, 1983; Strange, 1989) because acoustic effects of the gestures used in forming a vowel may be distributed over the entire periodic (transitions + target) portion of CVCs.

Much the same pattern seems to hold for nonnative speakers. Miranda and Strange (1989) presented words formed by inserting 10 native-produced English vowels into a /d_d/ context in a “full syllable” and a “silent center” condition, in which the center (target) portion of the vowel was removed. The rate of correct identifications was the same in the full-syllable and silent-center conditions for native English subjects (99%); it was slightly higher in the full-syllable than silent-center condition for nonnative subjects (63% vs. 58%). In a “center-only” condition, where the target remained and the transitions were removed, native speakers correctly identified 95% of vowels and nonnatives identified 74% of vowels.

Generally speaking, the fewer the response categories available to listeners, the higher will be the rate of correct identifications. As was shown earlier in Figure 2, listeners in the Eguchi and Hirsh (1969) study achieved
a higher correct identification rate with six than 12 response alternatives (71% vs. 46% correct). Verbrugge, Strange, and Shankweiler (1976) obtained a slightly lower correct identification rate for vowels in an /h_d/ context than did Peterson and Barney (1952). The difference (90% vs. 94%) may have been due to the fact that listeners in the Verbrugge et al (1976) study had 15 response alternatives (keywords such as heed, hid, hoed) whereas those in the Peterson and Barney study had just 10. Identification rates are generally lower when multiple talkers are presented than when the vowels of a single talker are presented in a constant-talker condition (Macchi, 1980; Strange et al, 1976). Thus another factor that may have contributed to the difference between the Verbrugge et al (1976) and Peterson and Barney (1952) studies was that a larger number of talkers were presented in each block in the former than the latter study (30 versus 10). Verbrugge et al (1976) presented /p_p/ words spoken by 15 talkers in two conditions to test whether listeners “calibrate” on the vowels spoken by a talker in a constant-talker condition. The overall rate of correct identifications was higher in a constant-talker than variable-talker condition (94% vs. 87%); the greatest improvement noted for any of the nine monophthongal vowels examined was for /æ/ (17%).

Other studies have shown that the time needed to recognize words increases, and the rate of correct identifications decreases (especially for words degraded by digital techniques), in a variable-talker compared to a constant-talker condition (Mullinex, Pisoni, and Martin, 1989; see also Assman et al, 1982). Recent results pertaining to /r/-/l/ identification by Japanese subjects suggest that the cognitive “cost” associated with processing different talkers in a variable-talker conditions, including a lower rate of correct identifications, may be greater for nonnative than native listeners (Logan et al, 1989).

Finally, nonnative talkers’ vowels may suffer more from non-ideal listening conditions than native speakers’ vowels if they are more distant from a native listener’s prototype than native speakers’ vowels. Nabeleck and Donahue (1984) found that the English spoken by 20 non-native speakers was as intelligible as that of 20 native speakers of English under normal listening conditions, but significantly less intelligible in reverberant conditions.
6. Dutch vs. English vowels

The present study examined Dutch subjects' production of six English vowels which, from the standpoint of Dutch, included vowels classifiable as identical, similar, or new. Four of the vowels defined the corners of the vowel quadrilateral (/i, æ, a, u/), one was a high front lax vowel (/I/), and the remaining vowel was a short, central vowel (/ʌ/).

Dutch is usually analyzed as having three diphthongs (/au/, /ʌ/, /e:i/) and 12 monophthongs, which Moulton (1962) divided into two classes based primarily on phonological considerations: short (/ɔ/, /ɑ/, /æ/, /I/, /ɛ/) or long (/i/, /u/, /y/, /o/, /a/, /ø/, /e/). Duration measurements have shown, however, that except before /r/, three of the supposedly long vowels (/i/, /y/, /u/) are short (Nooteboom, 1972; Nooteboom and Slis, 1972). In addition, Dutch has three long vowels that occur in foreign loanwords (/ɔ:/, /æ:/, /e:/) and a schwa (/ə/) in unstressed syllables.

Figure 3 shows the acoustic relationship between the six English vowels examined in the present study and the 12 Dutch monophthongs that occur in stressed syllables. The ellipses for English /i/, /I/, /æ/, /ʌ/, /a/, and /u/ represent 95% confidence intervals drawn around the F1-F2 values reported by Peterson and Barney (1952) for American English (henceforth AE) vowels as spoken by 33 adult male native speakers. The phonetic symbols indicating the mean F1-F2 values for Dutch vowels are based on values reported for 50 adult male native speakers of Dutch by Pols et al (1973). Figure 4 shows the relationship between the Pols et al. (1973) Dutch vowel data and the mean values reported by Holtse (1972) for English vowels as spoken by six male native speakers of British English (henceforth BE). The BE males were speakers of general Received Pronunciation.

6.1 An identical vowel

English /I/ was classified as an identical vowel because, as seen in Figures 3 and 4, the mean values for Dutch /I/ are very similar to those reported for the /I/ of AE and BE. An ANOVA carried out by Disner (1983) showed that the F1-F2 values for Dutch /I/ (Pols et al, 1973) and AE /I/ (Peterson and Barney, 1952) did not differ significantly. It appears that the /I/s of Dutch and BE do not differ auditorily, for Collins and Mees (1984: 86) stated that Dutch /I/ can “pass straight into (British) English without being modified” and should thus pose “no problem” for pronunciation.
Figure 3. The acoustic relationship between American English /i/, /I/, /a/, /æ/, /e/ and Dutch vowels. The ellipses for the English vowels surround approximately 95% of the F<sub>1</sub> and F<sub>2</sub> values reported by Peterson and Barney (1952) for 33 male English speakers. The phonetic symbols represent the mean F<sub>1</sub>-F<sub>2</sub> values reported by Pols, Tromp, and Plomp (1973) for 12 monophthongal Dutch vowels spoken by 50 native Dutch men.

Figure 4. The mean values for Dutch vowels (marked "D") reported for Dutch men by Pols et al (1969) and for English vowels (marked "E") reported for six British male speakers of RP by Holtse (1972).
6.3 A new vowel

English /æ/ was classified as new because Dutch has no vowel phoneme represented phonetically as /æ/. As can be seen in Figure 3, the mean values for Dutch /ɛ/ fall just within the upper boundary of the space occupied by English /æ/. Figure 4 shows that BE /æ/ is more distant from a Dutch vowel than any other BE vowel. Collins and Mees (1984) observed that both beginning and advanced Dutch learners of English substitute Dutch /ɛ/ for English /æ/ (see also van Heuven, 1986) but did not quantify their observation. Perceptual data reported by Schouten (1975) are consistent with the belief that Dutch learners of English will eventually recognize that English /æ/ is a new, non-Dutch vowel. Results were obtained from both advanced (3rd and 4th year) and beginning (1st year) Dutch students majoring in English at the University of Utrecht. The advanced students identified /æ/ more often than the beginning ones (93% vs. 57% correct). Not surprisingly, the only consistent error was /ɛ/ (12% of responses).

6.2 Similar vowels

English /i/ was classified as similar because Dutch /i/ is lower in the acoustic space than either the /i/ of AE (Figure 3) or BE (see Figures 3 and 4). Disner (1983) found that the F₁-F₃ values of Dutch /i/ and AE /i/ differed significantly. Consistent with this, Collins and Mees (1984) stated that Dutch learners typically substitute Dutch /i/ for English /i/ even though the Dutch vowel is "less fronted" than its English counterpart.

English /u/ was also classified as a similar vowel, but it may pose more of a learning problem for Dutch L2 learners than English /i/. Dutch /u/ differs acoustically from English /u/ (Disner, 1983) and has been described as intermediate to the /u/ and /U/ of English (van Heuven, 1986). Dutch learners may at first substitute Dutch /u/ for English /u/. More advanced Dutch learners are said to substitute the vowel "sequence" /yu/ for English /u/ (Collins and Mees, 1984), perhaps because /u/ is fronted so much in BE that some investigators regard it as a front vowel (Bauer, 1985). In a study by Schouten (1975), Dutch students identified synthetic English /u/ tokens correctly in only 60% of instances; most misidentifications of /u/ were as /U/.

English /ʌ/ would be classified as a new vowel if the only criterion applied was a consideration of phonetic symbols, for Dutch has no vowel phoneme represented as /ʌ/. However, several considerations led to /ʌ/’s
classification as similar. First, acoustic data suggest that English /a/ occupies a portion of the acoustic phonetic vowel space that goes unused in Dutch. As seen in Figure 3, the mean values for Dutch /a/ fall within the area occupied by AE /a/; and the same appears to be true for BE vowels. As seen in Figure 4, the BE /a/ is spectrally close to the short Dutch vowel /a/, and even slightly closer to the long Dutch /a/.19

There is anecdotal evidence that as Dutch speakers become more experienced in their L2, they continue to identify English /a/ with a vowel(s) of Dutch. van Heuven (1986) indicated that Dutch L2 Learners identify BE /a/ with the Dutch /œ/; and Collins and Mees (1984) indicated that even though the mid-high front rounded Dutch vowel /œ/ and BE /a/ are “very different” inexperienced Dutch learners of English substitute the Dutch /œ/ for English /a/ and experienced Dutch L2 learners tend to substitute the short Dutch vowel /a/ for /a/.20 Data reported by Schouten (1975) also suggests that interlingual identification continues. Dutch students identified /a/ correctly less often than /œ/ (41% of instances, as compared to 80% for synthetic /œ/s). There was only a small difference in the rates for /a/ obtained for advanced and beginning students (48% versus 32%).

Classification of the vowel in the word hot, represented here as /o/, is complicated by the fact that it is realized as a central, unrounded vowel (/a/) in AE but as a slightly rounded back vowel (/o/) in checked syllables in BE (Wells 1982). Figure 3 shows that AE /a/ is similar spectrally to the long Dutch vowel /a/, but tests performed by Disner (1983) indicated that these vowels had significantly different F₁ and F₃ (but not F₂) frequencies. Since AE /a/ is more distant from the short than the long Dutch /a/, it is likely that acoustic differences between Dutch /a/ and AE /a/ would also be significant.

It appears then, that from the standpoint of Dutch, AE /a/ should be regarded as a similar vowel. Figure 4 shows the acoustic relationship of /o/ — the BE equivalent to AE /a/ — to Dutch vowels. BE /o/ is close to Dutch /a/ in the F₁-F₂ space, but is considerably longer (see Holtse, 1972) than Dutch /a/, which Nooteboom (1973) suggests is represented centrally by Dutch speakers as being half the length of /a/. It is probably because of this temporal mismatch that even highly proficient Dutch L2 Learners continue to substitute Dutch /o/ for BE /o/ (Collins and Meese, 1984).21
9. Methods

9.1 Subjects

9.1.1 Talkers

The study examined English vowels spoken by 50 native Dutch talkers (25 males, 25 females) whose production and perception of English stops was examined in an earlier study (Flege and Eefting, 1987a). The Dutch talkers were individuals between the ages of 20 and 25 years who began learning English at about the age of 12 years in school. Forty of the students were majoring in English at the University of Utrecht, all of whom spoke English quite well. The remaining 10 students were majoring in engineering at a technical school in Delft. They had not studied English beyond the six years required in high school, and were consequently less proficient in English than the English majors.

Of the eight native speakers of BE selected to define the “norm” for English, four (all females) were recorded in Birmingham, Alabama and four (all males) were recorded in Amsterdam. None of these talkers could speak Dutch; all could be described as speaking General RP.

9.1.2 Listeners

The vowels were identified by three native speaker each of BE and AE. The responses of both BE and AE listeners were collected because dialect discrepancies between talkers and listeners may inflate the number of vowel identification errors (Macchi, 1980). AE can be heard in the Netherlands on television and in popular music, but BE is taught in the school (sometimes by native BE talkers). Consequently, Dutch speakers of English often seem to have a “British” accent to Americans.

The AE listeners (A1-A3) were native speakers of General AE. A3 was the author; A1 and A2 were part-time research assistants who participated in the experiment as part of their ordinary duties. The three BE listeners (B1-B3) were not speech researchers and were, accordingly, paid a nominal sum for participating. None of the listeners spoke Dutch or indicated familiarity with Dutch-accented English. None of the AE or BE listeners spoken with a marked regional accent.
9.2 Speech materials

The talkers produced test words formed by inserting six vowels (/i, I, α, u, a, æ/) into a single consonantal frame (/h_tI/), which yielded six real words: *heat, hit, hot, hoot, hut, and hat*. The words were read from a randomized list at the end of a the carrier phrase *Sip through a_.* The utterances were recorded using high-quality equipment (Sony Model TCD5M with Nakamichi CM-300 microphone). Three tokens of each word from the middle of the list were low-pass filtered at 5 kHz, digitized at a 12 kHz rate with 12-bit amplitude resolution, and stored on disk.

A waveform editor was used to discard all but the periodic portion (or "vowel", for short) from each stored waveform to avoid lexical bias effects in the subsequent identification experiment. Removing the noise associated with the initial /h/ and the stop gap and release burst for the final /t/ was considered preferable to having the talkers try produce isolated vowels because English lax vowels are not permitted to occur in open syllables. Moreover, prolonged, isolated vowels often have offglides not seen in CVCs. One disadvantage of the editing was that it was likely to decrease the overall rate of correct identifications, due in part to the loss of temporal information (see e.g., Bond, 1976).

9.3 Procedures for the intellibility test

9.3.1 Response categories

The listeners were told to identify the English vowels using 14 keywords listed on a response box in the following order: *heat, hoat*, *foot, hit, hat, hoot, hate, hot, hurt, hite, hout*, *het*, *hut, hoit*.* The four orthographic response categories marked by an asterisk are not real words in English, but all of the keywords conformed to English spelling conventions. Note that potentially confusable items such as *heat* and *hit* were not juxtaposed. The keywords represented all AE vowels except for /ɔ/ and /ɚ/. Schwa was excluded because it, like the Dutch schwa, does not occur in stressed syllables; and /ɔ/ was excluded because it did not contrast with /a/ in the dialect of the AE listeners.
9.3.2 Pre-Test

Owing to the large number of response categories, it was necessary to familiarize the listeners with the testing procedures before data collection began. The listeners always produced the keywords with the expected vowel when asked to read them (e.g., *hoit* was read as *[hoʰt]*)). They were given rhyming words for keywords that might be confused (e.g., *boy* for *hoit*). Next, the listeners were asked to point to keywords on the response box in response to the experimenter's live voice presentation of isolated vowels found in the 14 keywords.

Once the listeners were able to respond correctly and confidently to the isolated live-voice vowels, they participated in a formal pre-test. Isolated test vowels were prepared using the procedures described above. The pre-test consisted of the randomized presentation of three tokens of all 14 vowels (each edited from /hVT/ words). The vowels presented to AE listeners were edited from words read by a native speaker of AE; those presented to the BE listeners were from words spoken by a native speaker of BE. To be allowed to participate, a listener could make only two errors on the 42-item pre-test on the first and second days of testing (the errors could not be errors on the same vowel). One BE listener was unable to pass the pre-test and was replaced; another British listener passed the pre-test but was unable to complete the experiment, and so was also replaced.

9.3.3 Instructions

The listeners were told that they would hear vowels edited from /hVT/ words that had been spoken by an unspecified proportion of native speakers of Dutch and BE. Unlike the AE listeners, the BE listeners were unaware that the talkers had actually spoken only six /hVT/ words, not all 14 words listed on the response box. However, both the AE and the BE listeners were told to use all 14 response categories, as appropriate, and to guess if uncertain.

To signal their response, the listeners positioned the lever on the response box next to one of the 14 keywords. Each vowel was presented 1.0 sec after a response had been received for the preceding vowel.

9.3.4 Stimulus presentation

The 1044 vowel stimuli were normalized for overall RMS intensity and presented over headphones (TDH-49) at a comfortable level (76 dB SPL-A
peak syllable intensity). The 6 vowels x 3 tokens = 18 stimuli for each talker were randomly presented three times each in a single block. The order of the talkers was counterbalanced across listeners. About eight talkers were presented per day over an approximately two-week period. At least one native English talker was presented along with the Dutch talkers on each day of testing. This yielded a total of 3,132 responses from each listener (58 talkers x 6 vowels x 3 replicate tokens x 3 randomized presentations).

9.3.5 Evaluating degree of foreign accent

The experimental design called for grouping the Dutch talkers according to their pronunciation of English.

The foreign accent scores obtained for the Dutch talkers in a previous study (Flege and Eefting 1987a) were used for this purpose. In the earlier study, the 50 Dutch talker and five native speakers of BE produced sentences containing sounds known to be problematical for Dutch speakers of English: *I can read this for you; The good shoe fits Sue; The red book was good.* The sentences were digitized and randomly presented three times each to eight native speakers of BE residing in Birmingham, Alabama. In evaluating the 165 sentences, the listeners positioned a lever on a response box between endpoints marked “Strong Foreign Accent” and “No Foreign Accent”. The lever was connected to a linear potentiometer which, in turn, was connected to an 8-bit A-D converter that returned values ranging from 1 to 256 (that is, the higher the score the higher the degree of authenticity and the less the foreign accent).

The foreign accent scores obtained in this way averaged 239 (SD = 9) for the native BE talkers, 178 (SD = 36) for the 40 Dutch talkers majoring in English, and just 86 (SD = 29) for the 10 Dutch engineers. The Dutch talkers who received the 16 lowest scores were designated the “Strong Foreign Accent” group. The 16 Dutch talkers with the highest scores were designated the “Mild Foreign Accent” group even though two of them received scores in the native English range. The “Moderate Foreign Accent” group consisted of 16 talkers from the middle of the range.26

It is likely that the Dutch talkers with mild foreign accents had more English-language experience than those with strong accents. This assumption was supported by an analysis of stop consonant production. In Dutch, /t/ is implemented as a voiceless unaspirated stop with short-lag VOT values. Flege and Eefting (1987a) found that the 50 Dutch talkers’ foreign
accent scores correlated positively with the VOT measured in their productions of English /t/. The better their overall pronunciation of English, the more closely the Dutch students approximated the phonetic norm of English for /t/. Moreover, a language background questionnaire revealed that the engineering students, who were all rated as having strong accents, had less access to native speakers of English and spoke English far less frequently than the students majoring in English.

9.3.6 Analyses

The listeners' responses were stored on disk and later tabulated automatically. The dependent variable for each listener was the percentage of times, out of maximum of 9 responses (3 tokens x 3 presentations), that each of the six vowels spoken by the 58 talkers was identified as intended. Mean percent correct scores were then calculated for the three BE and three AE listeners. These scores, which were based on 27 forced-choice identification judgments for each vowel, were analyzed in mixed-design Group x Vowel ANOVAs. Separate analyses were carried out for the AE and BE listeners because only the AE listeners had speech training, and only they were aware of the vowels the talkers had identify of the six intended to produce.

10. Results of the intelligibility test

10.1 American listeners

The results are summarized separately for each of the three American listeners in Table 2. The mean percentage scores presented here are based on the frequency with which each of the six vowels was identified as intended. The phonetic symbols indicate error responses, that is, vowels heard instead of the intended vowels in more than five percent of instances.

Figure 5 (top) shows the mean percent correct scores for vowels spoken by the native BE speakers and the Dutch talkers with strong, moderate, and mild accents. The data shown here were averaged over the three AE listeners, who varied somewhat in the overall rate of correct identifications. Averaged across all six vowels, the correct identification rate was greater for native than nonnative talkers (94% versus 86%). Four vowels (/
Table 2. The mean rate at which three American listeners (A1-A3) identified vowels in six English words spoken by native speakers of British English (n = 8) and Dutch (n = 50). The numbers indicate the percentage of times each vowel was identified correctly. The phonetic symbols in parentheses indicate the vowels heard instead of the intended vowel in more than 5% of instances; their ordering indicates relative frequency of occurrence.

<table>
<thead>
<tr>
<th></th>
<th>British English talkers</th>
<th>Dutch talkers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>heat /i/</td>
<td>hit /u/</td>
</tr>
<tr>
<td>A1</td>
<td>99( )</td>
<td>99( )</td>
</tr>
<tr>
<td>A2</td>
<td>99( )</td>
<td>100( )</td>
</tr>
<tr>
<td>A3</td>
<td>99( )</td>
<td>100( )</td>
</tr>
<tr>
<td>M</td>
<td>99( )</td>
<td>100( )</td>
</tr>
<tr>
<td></td>
<td>heat /i/</td>
<td>hit /u/</td>
</tr>
<tr>
<td>A1</td>
<td>99( )</td>
<td>100( )</td>
</tr>
<tr>
<td>A2</td>
<td>100( )</td>
<td>100( )</td>
</tr>
<tr>
<td>A3</td>
<td>98( )</td>
<td>100( )</td>
</tr>
<tr>
<td>M</td>
<td>99( )</td>
<td>100( )</td>
</tr>
</tbody>
</table>

The native BE talkers’ /æ/ and /ʌ/ were produced by the native BE talkers were identified at near-perfect rates, but two of their vowels (/æ/, /ʌ/) were identified in less than 90% of instances.

The native BE talkers’ /æ/ and /ʌ/ were also identified at near-perfect rates, and their /æ/ and /ʌ/ tended to be misheard as /a/. The relatively poor intelligibility of /æ/ and /ʌ/ as spoken by the BE talkers agrees with results obtained previously for AE talkers’ production of these vowels (see Table 1). However, the AE listeners identified the vowel in *hot*, which is implemented as /u/ in BE, somewhat better than the vowel that Americans produce in this word (viz. /a/) has been identified previous vowel identification experiments.

As for the native BE talkers, The Dutch talkers’ /i/ and /I/ were also identified at near-perfect rates, and their /æ/ and /ʌ/ were identified relatively poorly. The greatest difference between the native and Dutch was for /u/ and /ɔ/ because their /u/ was heard as /U/ and their /ɔ/ (/o/) was heard as /ʌ/. Other common substitutions heard in the Dutch-accented English were /ʌ/-for-/æ/ and /ɛ/-for-/æ/, and /ɔ/-for-/ʌ/. The only important difference in substitution errors heard for the native and Dutch talkers was
Figure 5. (top) The mean rate at which American English listeners correctly identified six English vowels (/i/, /I/, /u/, /o/, /æ/, /æ/) spoken by Dutch talkers who spoke English with relatively strong, moderate, or mild foreign accents (16 per group) and by native speakers of British English (n = 8); (bottom) the results obtained for three British English listeners. The error bars enclose +/− one standard error.
that the Dutch talkers' /æ/ attempts were sometimes heard as /ɛ/. The /ɛ/ variants were likely to have been the result of cross-language phonetic interference, for /ɛ/ seems to be the Dutch vowel that is closest to English /æ/.

Figure 5 (top) shows that there were systematic differences between the three Dutch groups for certain vowels but not others. Between-group differences did not exist for either /i/ or /I/ because these vowels were identified consistently at near-perfect rates. The rate for /I/ was about 80% correct for all three Dutch groups. For /aː/ and /æ/, the percent correct scores increased as a function of how well the Dutch talkers spoke English. For /aː/, the Dutch talkers with strong, moderate, and mild foreign accents had percent correct scores of 70%, 87%, and 90%, respectively. For /æ/, the percent correct scores were 47%, 74%, and 80%. Conversely, the scores for /u/ seemed to decrease as a function of how well the Dutch subjects pronounced English. The percent correct scores for the Dutch talkers with relatively strong, moderate, and mild foreign accents were 94%, 87%, and 81%, respectively.

To determine which, if any, of the three Dutch groups differed from the native BE talkers, the percent correct scores obtained for the AE listeners were submitted to a mixed-design ANOVA. The significant Group x Vowel interaction obtained \[ F(15,260) = 2.64, P < 0.001 \] was followed by tests of simple main effects, which suggested that learning had taken place for two of the six English vowels. The Group effect was significant for /o/ \[ F(3,52) = 4.09, P = 0.011 \] and /æ/ \[ F(3,52) = 4.48, P = 0.007 \] but was non-significant for the other four vowels examined (p > 0.10). Post-hoc tests (Newman-Keuls, \( \alpha = 0.05 \)) revealed that for both /æ/ and /o/, the Dutch talkers with moderate and mild foreign accents had higher correct identification rates than those with strong accents but these two groups did not differ from the native BE talkers.

The present experiment was undertaken to test the speech learning model outlined in earlier sections. In this context it is important to note that the results for the new vowel /æ/ were predicted by the SLM, but not the results for the similar vowel /o/.\(^{30}\)

10.2 British English listeners

The misidentifications of the three BE listeners are summarized in Table 3. The mean rates at which these listeners identified vowels spoken by native
Table 3. The mean rate at which three native British English listeners (B1-B3) identified vowels in English words spoken by native speakers of British English (n = 8) and Dutch (n = 50). The numbers indicate the percentage of times each vowel was identified correctly. The phonetic symbols in parentheses indicate the vowels heard instead of the intended vowel in more than 5% of instances; their ordering indicates relative frequency of occurrence.

<table>
<thead>
<tr>
<th>Heat /i/</th>
<th>Hit /i/</th>
<th>Hoot /u/</th>
<th>Hot /aa/</th>
<th>Hat /æ/</th>
<th>Hut /ʌ/</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1 85( )</td>
<td>82(ε,3)</td>
<td>82(U,o)</td>
<td>82(ʌ)</td>
<td>86(ɑ,ʌ)</td>
<td>76(ɑ,ʌ)</td>
<td>82</td>
</tr>
<tr>
<td>B2 95( )</td>
<td>100( )</td>
<td>99( )</td>
<td>96( )</td>
<td>57(ʌ,ɛ)</td>
<td>85(ɑ)</td>
<td>89</td>
</tr>
<tr>
<td>B3 93( )</td>
<td>100( )</td>
<td>100( )</td>
<td>85(ʌ)</td>
<td>75(ɛ,ɑ,ʌ)</td>
<td>72(ɑ,ɛ)</td>
<td>88</td>
</tr>
<tr>
<td>M 91( )</td>
<td>94( )</td>
<td>94( )</td>
<td>88(ʌ)</td>
<td>73(ʌ,ʌ)</td>
<td>78(ɑ,ɛ,ɑ,ʌ)</td>
<td>86</td>
</tr>
</tbody>
</table>

British English talkers

<table>
<thead>
<tr>
<th>Heat /i/</th>
<th>Hit /i/</th>
<th>Hoot /u/</th>
<th>Hot /aa/</th>
<th>Hat /æ/</th>
<th>Hut /ʌ/</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1 94( )</td>
<td>83(ε)</td>
<td>66(U,o)</td>
<td>73(ɑ,ʌ)</td>
<td>80(ɑ,ʌ)</td>
<td>55(ɑ,ʌ)</td>
<td>75</td>
</tr>
<tr>
<td>B2 99( )</td>
<td>96( )</td>
<td>82(U,o)</td>
<td>78(ʌ)</td>
<td>60(ɛ,ʌ)</td>
<td>71(ɑ,ɑ,æ)</td>
<td>81</td>
</tr>
<tr>
<td>B3 97( )</td>
<td>100( )</td>
<td>85(U)</td>
<td>86(ʌ)</td>
<td>76(ɛ,ʌ)</td>
<td>62(ɛ,ɑ)</td>
<td>84</td>
</tr>
<tr>
<td>M 97( )</td>
<td>93( )</td>
<td>78(U,o)</td>
<td>79(ʌ)</td>
<td>72(ɛ,ʌ)</td>
<td>63(ɑ,ɑ,ʌ)</td>
<td>80</td>
</tr>
</tbody>
</table>

Dutch talkers

BE talkers and talkers in the three Dutch groups are shown in Figure 3 (bottom).

The BE listeners were expected to be better able than the AE listeners to identify vowels spoken by their fellow countrymen, but the rates of correct identification of vowels was actually lower for the BE than AE listeners (86% vs. 94% correct). The most likely explanation for this paradoxical finding is a methodological difference. Recall that the AE but not the BE listeners were aware that the talkers had actually intended to produce just six vowels, not all 14 vowels shown on the response box.

In support of this explanation, Table 3 reveals that the BE listeners used a wider range of response variants in identifying vowels than the AE listeners. One response variant used by the BE but not the AE listeners in identifying /ɑ/ (/ɑ/ in AE 31) was /ʌ/. In addition to hearing /ʌ/-for-/æ/ substitutions, the BE listeners also heard /ɛ/-for-/æ/ substitutions. And in addition to hearing /ɑ/ (/ɑ/-for-/ʌ/, the BE listeners also heard /æ/-for-/ʌ/. In general, however, the misidentifications of the BE and AE listeners were similar. The BE listeners misheard /æ/ as /ʌ/. They heard /ʌ/ as either /ɑ/ or /æ/, /u/ as /U/, and /o/ as /ʌ/.

The BE listeners also identified fewer vowels spoken by the Dutch talkers than the AE listeners (80% vs. 86%). As shown in Figure 5 (bottom),
they identified the Dutch talkers' attempts at /i/ and /I/ most of the time (97% and 93% correct, respectively). The BE listeners identified the Dutch talkers' /u/s and /o/s less often (78%, 79%) than the Dutch talkers' high front vowels, and identified their /æ/ and /ʌ/ attempts even more poorly (72%, 63% correct).

There were between-group differences in intelligibility for all six vowels. Somewhat surprisingly, the /i/s produced by the Dutch talkers in all three groups were uniformly higher than the native BE talkers' /i/s. There was a small increase in the rate of correct /I/ identifications as a function of accent. The rates for the Dutch subjects with strong, moderate, and mild accent were 90%, 94%, and 96%, respectively. A similar pattern was evident for /o/ (72%, 78%, and 86%), /æ/ (50%, 84%, 82%) and /ʌ/ (56%, 68%, 65%). For /u/, on the other hand, performance seemed to deteriorate slightly along with improvements in accent (82%, 77%, 74%).

As expected, the ANOVA yielded a significant Group x Vowel interaction \[F(15,260) = 2.34, p = 0.004\]. The simple main effect of Group was significant for /æ/ \[F(3,52) = 6.42, p < 0.001\] but non-significant for all other vowels, including /o/. Post-hoc tests revealed that the Dutch talkers with moderate and mild foreign accents had higher scores for /æ/ than those with strong accents \((p < 0.05)\). The moderately and mildly-accented Dutch talkers did not differ significantly from the native BE talkers.

These results supported the prediction of the SLM for /æ/. It is important to note that, for the BE listeners, the Dutch subjects did not show a significant improvement in /o/ intelligibility (the result obtained earlier for AE listeners).

To understand better why the AE listeners but not the BE listeners registered a significant improvement for /o/, which was classified as a similar vowel, the variants that the BE and AE listeners gave in identifying the vowel in \textit{hot} as spoken by mildly and strongly-accented Dutch talkers were tabulated. The AE listeners used fewer error variants, and applied those variants to fewer Dutch talkers in the Mild than Strong Foreign Accent group. Unlike the AE listeners, the BE listeners did not give fewer error variants for talkers in the Mild than Strong Foreign Accent groups, but they did show a decrease in the number of talkers to whom they applied the error variants they used.
Figure 6. The mean rate of correct identifications of /æ/ as spoken by Dutch talkers who spoke English with strong, moderate, or mild foreign accents (16 per group) and eight native speakers of British English ("NE"). The top panel shows results obtained from native American English listeners; the bottom panel shows results obtained for British English listeners. Each data point is based on 27 judgments (3 tokens x 3 presentations x 3 listeners).
10.3 Individual talkers' production of the new vowel /æ/.

The SLM generated the prediction that Late L2 learners will produce a new L2 vowel authentically once they recognize it to be new. Since little is known at present concerning how long such recognition may take, Figure 6 was prepared to show the correct identification rates for the /æ/ spoken by 48 individual Dutch talkers.

Figure 6 (top) shows the rates obtained from the AE listeners; and Figure 6 (bottom) shows the rates obtained for the BE listeners. Note that some talkers in all three Dutch subgroups produced the new vowel /æ/ well, and some produced it poorly. The AE listeners seldom identified one native BE talker's /æ/ correctly, frequently hearing his intended /æ/ tokens as /aU/ or /ə/. This talker was a 25-year-old male from Preston, England who was living in Amsterdam. According to James (1989), this individual had been exposed to a “northwest regional influence (i.e., Lancashire)” which might be expected to have caused a backing and/or lowering of his /æ/. The range of correct identifications for the remaining BE talkers was 78%-100% correct.

All three subgroups of Dutch talkers showed a wide range for /æ/. Just three of the 16 strongly accented Dutch talkers had rates exceeding 75% correct, whereas 11 moderately accented and 12 mildly accented Dutch talkers had rates above this level.

The range of percent correct scores for the native BE talkers was wider in the data obtained from the BE than the AE listeners. The one native BE talker whose /æ/ was poorly identified by the AE listeners was also poorly identified by the BE listeners (22% correct for both listener groups). However, several talkers' /æ/ were identified less often by the BE than the AE listeners. One talker's identification rate for /æ/ dropped from 100% to 40% correct because the BE listeners heard his /æ/ attempts as /ɛ/; another talker's /æ/ rate dropped from 96% to 67% because the BE listeners heard /ʌ/.

11. Acoustic measurements

The intelligibility study supported the SLM's prediction that a new vowel (English /æ/) would be produced authentically by at least the most proficient Dutch speakers of English. However, it provided little support for the
model’s prediction concerning English vowels classified as “similar” (viz. (/i/, /u/, /ɔ/, /o/)). It was predicted that the Dutch talkers would produce the similar English vowels in a Dutch-like manner; Since the similar Dutch vowels diverged from the phonetic norm of English for corresponding English vowels, they were expected to be less intelligible when produced by even by highly proficient Dutch speakers of English than by native BE talkers. However, this prediction was not supported.

The possibility exists that persistent differences between native and non-native speakers did exist for the similar vowels, as predicted, but that the intelligibility test did not have a sufficiently fine resolution to reveal the differences. After all, the intelligibility scores were based on the percentage of correct identifications, and identification implies the reduction (or elimination) of fine-grained stimulus characteristics. An acoustic analysis was therefore carried out to further test the prediction of the SLM concerning similar vowels.

11.1 Procedures

Fourteen linear prediction coefficients were computed to estimate the center frequency of the first two formants ($F_1$, $F_2$) in vowels spoken by the 10 Dutch males with the best accents and the 10 Dutch males with the poorest accents. For the LPC analysis, a 25.6-ms Hamming window was placed at the acoustic midpoint of the vowels in *heat*, *hit*, *hat*, *hot*, *hut*, and *hoot*.

To provide a sufficiently large number of BE talkers to permit statistical comparisons of native and Dutch talkers, the data reported by Holtse (1972) for six male native speakers of standard RP was added to data obtained for the four BE males who participated in the present study. Holtse measured formants in the vowels found in *heat*-heed, *hit*-hid, *hat*-had, *cut*-cud, *heart*-hard, and *coot*-cooed as spoken by six male native speakers of RP. Most of the subjects were students or faculty members at the University of Copenhagen. Holte’s (1972) study differed from the present one in several respects. Vowels were measured spectographically rather than by LPC analysis; means were based on a larger number of tokens (viz. 10-16); and some of the lexical items examined in the two studies differed. However, the data for the four BE males from this study did not appear to differ systemmatically from Holtse’s six BE male subjects.
11.2 Results and discussion of the acoustic analysis

The mean values (in Hz) for the four groups of male talkers are presented in Table 4. The mean values are plotted in a Mel scale in Figure 7, which shows that the Dutch subjects with a relatively good English pronunciation more closely approximated the native BE talkers’ vowel system than the Dutch subjects with relatively poor accents. Most of the difference between the native and Dutch speakers, and between the two Dutch groups, was confined to the new vowel /æ/ and the similar vowels /ʌ/ and /ɔ/ (/a:). There seemed to be little difference between groups for the two similar vowels /i/ and /u/.

To determine which, if any, of these differences was significant, the 30 mean $F_1$ values (3 groups x 10 subjects) for each of the six vowels were submitted to separate ANOVAs, as were the mean $F_2$ values. The alpha level was set at 0.008 to obtain a per-experiment error rate of 0.05 for the $F_1$ and $F_2$ tests. The difference between groups was non-significant for the new vowel /æ/. As expected from previous cross-language research (Disner, 1983), the between-group differences for the identical vowel /I/ were non-significant. It is noteworthy, however, that the magnitude of the non-significant acoustic differences between groups for this vowel were about the same as those for the similar vowels /i/ and /u/.

Table 4. The mean first and second formants frequencies (in Hz) for six English vowels spoken by 10 male native speakers of British English (4 from the present study and 6 from Holtse, 1972), 10 native Dutch males with good accents in English, and 10 native Dutch males with relatively poor (i.e., strong) accents in English.

<table>
<thead>
<tr>
<th>Native speakers of English</th>
<th>/i/</th>
<th>/I/</th>
<th>/æ/</th>
<th>/ʌ/</th>
<th>/ɔ/</th>
<th>/u/</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1$</td>
<td>270(56)</td>
<td>405(43)</td>
<td>715(93)</td>
<td>681(85)</td>
<td>649(103)</td>
<td>294(40)</td>
</tr>
<tr>
<td>$F_2$</td>
<td>2410(74)</td>
<td>2033(141)</td>
<td>1644(178)</td>
<td>1294(101)</td>
<td>1021(113)</td>
<td>1232(236)</td>
</tr>
</tbody>
</table>

Dutch with, good accents in English

<table>
<thead>
<tr>
<th>/i/</th>
<th>/I/</th>
<th>/æ/</th>
<th>/ʌ/</th>
<th>/ɔ/</th>
<th>/u/</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1$</td>
<td>283(29)</td>
<td>390(21)</td>
<td>638(85)</td>
<td>572(89)</td>
<td>526(50)</td>
</tr>
<tr>
<td>$F_2$</td>
<td>2255(186)</td>
<td>1895(129)</td>
<td>1482(96)</td>
<td>1217(57)</td>
<td>961(85)</td>
</tr>
</tbody>
</table>

Dutch with poor accents in English

<table>
<thead>
<tr>
<th>/i/</th>
<th>/I/</th>
<th>/æ/</th>
<th>/ʌ/</th>
<th>/ɔ/</th>
<th>/u/</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1$</td>
<td>281(33)</td>
<td>389(33)</td>
<td>604(186)</td>
<td>531(125)</td>
<td>516(83)</td>
</tr>
<tr>
<td>$F_2$</td>
<td>2215(227)</td>
<td>1884(132)</td>
<td>1615(131)</td>
<td>1296(91)</td>
<td>1018(120)</td>
</tr>
</tbody>
</table>
As predicted by the SLM, a between-group difference was obtained for the $F_1$ values in the similar vowels /\ and /ν/ [$F(2,27) = 5.86, 8.24$]. Newman-Keuls post-hoc tests revealed that the native BE talkers' $F_1$ frequencies were significantly higher than those of both Dutch groups. The differences for the other two similar vowels (/i/, /u/) were non-significant, however, as were all of the between-group differences for $F_2$.

The acoustic results partially confirmed the SLM's prediction that similar vowels will represent a persistent pronunciation problem for late L2 learners. One reason why between-group differences were not seen for two of the four similar vowels may have been that the acoustic differences between the two similar English vowels and their counterparts in Dutch were too small to detect reliably.
12. General discussion

The present study tested a speech learning model (SLM) by examining the intelligibility of identical, similar, and new English vowels spoken by native speakers of British English (BE) and Dutch talkers who spoke English with relatively strong, moderate, and mild foreign accents. As expected from previous vowel intelligibility studies (see Table 1), the high English vowels /i/, /I/, and /u/ were identified more often than the non-high vowels /æ/, /æ/, and /ʌ/. The between-vowel differences were not significant for the native BE talkers, either when their vowels were assessed auditorily by fellow native speakers of British English or by native speakers of American English.

It came as no surprise that English /I/ was as intelligible when spoken by talkers in all three Dutch subgroups as when spoken by the native BE talkers. This vowel was classified as “identical” to the /I/ of Dutch because previous research had shown little or no difference between the /I/ of Dutch and that of English. A different pattern of results was obtained for English vowels classified as new and similar, however.

12.1 The production of new vowels

Important between-vowel differences existed for Dutch talkers with mild, moderate and strong foreign accents. The SLM predicted that in early stages of learning, Late L2 learners may initially fail to produce new L2 sounds authentically, but will eventually do so as a result of establishing phonetic categories for them. In support of this, the Dutch talkers with moderate and mild foreign accents produced /æ/ significantly more intelligibility than Dutch talkers with strong foreign accents. The intelligibility of English /æ/ as spoken by the moderately and mildly-accented Dutch talkers did not differ significantly from that for the native /æ/s spoken by BE talkers. The Dutch talkers with strong accents produced /æ/ significantly less intelligibly than /I/, /i/, or /u/ whereas the intelligibility of the /æ/s produced by the Dutch talkers with moderate and mild foreign accents did not differ from that of the high vowels.

The result for /æ/ agrees with the finding by Bohn and Flege (1991) that Germans are able to learn English /æ/. Taken together with that finding, the results presented here show that phonetic learning remains possible beyond early childhood, and support an important prediction of the SLM,
namely that L2 vowels occupying a portion of the phonetic space not exploited by the L1 vowel system can be “learned”, that is, produced authentically, by Late L2 Learners.

There are three reasons for caution, however. The first is that Dutch has English loanwords with /æ/ such as *fan* and *dancing*, and that the vowels in these words may be realized with a slightly lower quality than the Dutch /e/ (James, 1989). One might argue, then, that /æ/ wasn’t really a new vowel.

A second reason for caution is that the Dutch talkers who succeeded in producing /æ/ authentically had first begun learning English in school at the age of 12 years. One might argue that they had not yet passed a “critical period” and that they should not be classified as late L2 learners. However, data reported by Flege and Fletcher (1991) suggests that the offset of the critical (or sensitive) period for human speech learning occurs long before the age of 12 years.

The third reason for caution is that the Dutch subjects who succeeded in producing English /æ/ s authentically were all students majoring in English at the University of Utrecht. Perhaps they succeeded because of some special training or talent. If so, then the results may not generalize to other L2 learners who are equally experienced in English. Many Dutch L2 learners examined here produced English /æ/ s that were identified in 100% of instances by both American (AE) and British English (BE) listeners but, contrary to the SLM’s prediction, some talkers in all three Dutch groups—including the Mild Accent group made up exclusively of students majoring in English — produced /æ/ poorly. This leads one to ask “Why did some Dutch subjects but not others succeed in learning /æ/?”

12.2 The production of similar vowels

An hypothesis of the SLM is that category formation for similar L2 sounds is blocked by equivalence classification. This generates the prediction that Late L2 Learners will phonetically approximate the phonetic norm for similar L2 vowels but will never produce such vowels authentically.

In agreement with this, there was not a significant difference in intelligibility between the Dutch talkers with strong, moderate, and mild foreign accents for three of the similar L2 vowels examined (/ɪ/, /u/, and /ʌ/). However, contrary to hypothesis, the Dutch groups did not differ from native speakers of English for these vowels; and improvement with L2 experience
was noted for the similar vowel /ɒ/. Intelligibility for /ɒ/ was higher for the moderately and mildly accented Dutch talkers than for the strongly accented Dutch talkers.

One might argue that the improvement for /ɒ/ did not provide counterevidence to the SLM’s prediction about similar vowels because /ɒ/ is used to produce words like *hot* in British English whereas /ʌ/ is the most common vowel used in such words in American English. An improvement in the Dutch talkers’ /ɒ/ production was noted in the data obtained for British English (BE) but not American English (AE) listeners. However, the lack of a difference between the Dutch and native speakers for the similar vowel /ɒ/ represents a more serious challenge to the model.

The SLM hypothesizes that similar L2 vowels pose a persistent problem for Late L2 Learners because they continue to be identified with vowels in the L1 as the result of equivalence classification. This led to the prediction that similar English vowels spoken by the Dutch subjects, even those with very good overall accents, would differ from those produced by native speakers of English. More specifically, it was predicted that the Dutch talkers would show “compromise” acoustic values representing an assimilation of the phonetic properties of the corresponding L1 and L2 vowels that were equated at a phonetic category level.

The present study failed to show a significant difference in intelligibility between native and Dutch talkers for any of the similar vowels examined, but an acoustic analysis supported partially the prediction concerning similar vowels. It revealed significant formant frequency differences in the production of the similar vowels /ɒ/ and /ʌ/ between native speakers and Dutch L2 learners, both those with good and relatively poor accents. These two similar English vowels were probably more distant from their closest Dutch counterparts than the other two similar English vowels (viz. /i/, /u/) that did not yield a native versus nonnative difference.

Even if there were a difference between the native and nonnative speakers’ /i/’s and /u/’s, as predicted, acoustic measurement resolution may have been too coarse-grained to show it. That is, the similar English vowels /i/ and /u/ may be so close to the corresponding Dutch vowels that even the unmodified substitution of a Dutch vowel for a similar English vowel might not reduce intelligibility or be detectable in acoustic analysis. The possibility exists that a more fine-grained auditory analysis, such as the paired comparison of Dutch and English /i/’s (and /u/’s) in an accent-detection test (see Flege, 1984), would support the prediction of a continued difference
between native and nonnative speakers for the similar L2 vowels. Whatever
the outcome of such a test, it is clear that the larger the L1 vs. L2 difference
in similar vowels, the easier they are to measure acoustically (and probably
to detect auditorily).

One trend seen for both the BE and AE listeners is worthy of further
comment. According to Collins and Mees (1984), the production of English
/u/ may deteriorate as Dutch speakers gain experience in English. They
stated that experienced Dutch speakers of English begin to substitute /yu/
for English /u/. The present study showed that percent correct scores for /u/
decreased non-significantly as foreign accents improved. Perhaps the Dutch
talkers with the best English pronunciations had begun reproducing the
very fronted, [y]-like realizations of /u/ used by native speakers of BE
(Bauer, 1985). If so, the vowel of boot may have been misidentified by
native BE and AE listeners when presented in isolation. A question for
future study is whether an apparent deterioration is the production of Eng­
lish /u/ would be noted if vowel tokens were presented in context (see Sec­
tion 5).

This chapter provided a more explicit operational definition of the dis­
tinction between similar and new vowels than has been offered in earlier
work still, an important question that remains to be answered is “How
much must an L2 vowel differ from vowels in the L1 to be regarded as
new?” It remains to be determined if there is an absolute phonetic differ­
ence threshold that, once crossed, triggers the formation of a new phonetic
category. Future research may show that the new vs. similar vowel distinc­
tion, can be defined only on the basis of an individual learner’s vowel sys­
tem, not generalizations drawn for a whole population (i.e. a language).
Patterns of inter-lingual identification may also be influenced by as-yet
undefined individual modes of phonetic processing.

Acknowledgments

This research was supported by NIH grant DC00257. The author expresses gratitude to
the colleagues who recorded subjects in The Netherlands (W. Eefting, A. James) and
provided information about Dutch (B. Elsendoorn, V. van Heuven), the research assis­
tants who helped with various phases of the project (B. Hartley, L. Skelton, and H.
Slocom), and especially the British English listeners who made so many visits to the
Bicommunication Research Laboratory for testing. Thanks are also extended to O.-S.
Bohn for comments on a previous draft of this chapter.
Notes

1. One might question whether the English subjects had to learn something in order to produce French stops authentically for it is likely that, as adults, they produced English /d/ with short-lag VOT and, as children, they implemented English /t/ with short-lag VOT values (Zlatin and Koenigsknecht, 1976).

2. The earliest version of the SLM (Flege, 1981) was called the "phonological translation" hypothesis because it was then believed that interlingual identification occurred at a phonemic level. The model's name has been changed owing to the current belief that interlingual identification occurs at a phonetic level of analysis (see below). Another change made since the first statement of the model is that age of learning is now regarded as a crucial factor in determining how authentically certain L2 sounds will be produced.

3. It appears that not all phonetic contrasts will be lost in the absence of environmental stimulation. Best et al. (1988) found that infants being reared in an English-speaking environment remained able to discriminate Zulu click contrasts even though clicks are not found in English. These authors hypothesized that both the acoustic nature of a novel phonetic contrast and its relation to phonetic categories and contrasts in the language spoken around an infant play a role in determining which phonetic contrasts will be lost.

4. An alternative to the use of prototypes in a developmentally-oriented model of speech perception was recently described by Jusczyk (1989). Prototype models assume the existence of an abstract, underlying representation to which incoming phones are matched. During speech learning, the prototype is refined by adding additional detail and/or by re-specifying the weighting of various properties. According to a multiple-trace model, on the other hand, phones are identified by comparison to multiple individual instances. Experience leads to the addition of individual traces, which are based on exemplars used collectively in recognition. Such an approach might provide a perceptual basis for the different phonetic realization rules Late L2 learners use to distinguish similar L1 and L2 sounds in the absence of different phonetic categories (see below).

5. It appears that the "tuning" hypothesized by Aslin and Pisoni (1980) referred to experience-based modifications of phonemic categories. Best et al. (1988) also hypothesize that the detection of divergences from the phonetic norms of a language occurs at a phonemic level.

6. This procedure for estimating a perceptual norm for L2 vowels must be regarded as a first approximation, and must be used with caution. It is conceivable, for example, that talkers could produce a vowel in a way that differs from their internal perceptual representation by using a realization rule. This caution is especially applicable to production data gathered in a formal experiment, where maximally clear, careful speech is likely to be elicited.

7. The difference in symbols is by no means trivial, for it leads to different expectations concerning L2 learning. For example, symbolizing the lax vowel as /I/ suggests that it will be regarded as a new vowel by Spanish learners of English whereas symbolizing it as a short /i/ suggests that it will be regarded as a similar vowel.

8. Direct magnitude estimation might be used to estimate the phonetic similarity of vowels in L1 and L2. The potential problem with this method is that listeners might evaluate dis-
stances using an auditory metric (Pols et al., 1969), a phonetic metric, or some combination of both (see Kewley-Port and Atal, 1989).

9. New sounds might require a longer time to identify than similar sounds in a speeded classification test. Similar sounds might be easier to remember than new vowels in a list learning experiment.

10. These predictions concerning the production of similar L2 stop consonants are founded on two important assumptions: additional realization rules with which to produce similar L2 stops can be established even though additional categories cannot; and substantial VOT differences between L1 and L2 stops cannot be achieved through the use of realization rules.

11. It should be noted, however, that the equally experienced French learners of English showed only a non-significant difference in the expected direction from English monolinguals in producing English /u/. The seemingly divergent results for the French and English subjects was surprising since the subjects in both groups had lived in an L2-speaking environment for about 12 years on the average.

12. Portuguese does not have a phoneme whose most characteristic allophone is [e]. Portuguese has an /e/ and /e/. Acoustic data (Godinez, 1978) suggest that Portuguese /e/ is produced with a wide range of variants. It overlaps the Portuguese /e/ category in an F1-F2 space. Portuguese /e/ seems to occupy a portion of the acoustic space occupied by English /æ/, but an even larger portion of the space occupied by English /e/.

13. The finding just reported is based on my own analysis of data reported by Major (1987), which were obtained at the Biocommunication Research Laboratory.

14. Interestingly, the /æ/s and /i/s (and also /l/s) produced by a wide range of text-to-speech devices are also misidentified at relatively higher rates (Logan et al., 1989).

15. Such variation also seems to exist in L2 speech production. Ma and Herasimchuk (1971, cited by Amastae, 1978) found that English /æ/ was produced correctly by Puerto Ricans in only about half of instances. The expected Spanish “interference” variant (/a/) was used in about 30% of instances, and an /e/ “approximation” variant was used in about 20% of instances. The interference variant was more frequent in the most formal speaking context than in less formal contexts.

16. A high correct identification rate can be obtained under certain circumstances, however, even when the listener is offered as many as 15 response alternatives. Verbrugge et al. (1976) cite a study by Abramson and Cooper (1959) which yielded a 96.0% identification rate with 15 response alternative. The vowel tokens in that study were carefully selected, the listeners were personally familiar with the talkers, there were relatively few talkers (8), and the total number of vowels produced by each talker was relatively large (15).

17. The ellipses are drawn through four points located two standard deviations from the intersection of axes defining the two principal components of variation.

18. Elsendoorn (1986) offered the same view. Data reported by Schouten (1975) supports the belief that L2 learners may not recognize immediately the identity of L1 and L2 vowels.
In that study, 77 native Dutch students majoring in English at the University of Utrecht identified synthetic vowels with a fixed 175-ms duration using 15 phonetic symbols and keywords. The synthetic English /I/ were identified correctly in only 71% of instances, probably because of difficulty using phonetic symbols, or because the fixed duration was more appropriate for Dutch /e/ than /I/ (van Heuven, 1986). The /I/ were misidentified as /e/ in 13% of instances, and as /e/ in 12% of instances.

19. It should be noted, however, that a great deal of variation exists in the production of BE /I/ (Collins and Meese, 1984). Barry (1974) indicated that there are two distinct “norms” for /I/, one for RP English and another, more centralized regional English variant. This observation was later confirmed by a study comparing acoustic values obtained in a number of studies examining English /I/ (Bauer, 1985).

20. Holtse (1972, p. 9) noted the presence of additional low amplitude formants in BE /I/ which, although not classified as F₁ or F₂, might cause this vowel to be perceived as “closer in quality” than might be suggested by its position in an F₁-F₂ space. According to Elsendoorn (1986), BE /I/ is perceptually close to the Dutch schwa but would probably not be identified with it because the Dutch schwa occurs only in unstressed syllables.


22. Of these subjects, one was born in Glasgow, Scotland; one in Preston, England; and one in London (for whom biographical data is not available). These talkers were described (James, 1985b) as speaking a form of English free of any of “class-defining” or “regional accent” features. The female talkers all held advanced degrees. Their birthplaces were Oxford, Darley, Liverpool, and Kenya (Africa).

23. Listener A1 was a female, aged 31, who was born and raised in San Francisco; A2 was a female, aged 26, from the Washington, D.C.; A3 was a male, aged 37, from Cincinnati, Ohio. Listeners A1 and A2 were newly arrived in Birmingham, whereas A3 had lived there for six years, and had lived for two years in Western Europe.

24. Listener B1 was a 39-year-old male scientist at the University of Alabama at Birmingham who had arrived in Birmingham just prior to the experiment. B1 was born in London and had lived in the London area (Hertfordshire, St. Albans) up to the age of 18 years. He had lived in a variety of places in Great Britain (Liverpool-3 years, Cambridge-3 years, Leicester-3 years) and had completed a two-year post-doctoral fellowship in Raleigh, NC. B1 characterized his speech as a “London” accent. Listener B2 was a 30-year-old male UAB scientist who had lived in Bristol until the age of 18 years. After three years of university work at Cambridge, he spent the following six years earning a Ph.D. in Biochemistry at the State University of New York at Stoneybrook. Listener A3 was an 18-year-old male Londoner who was spending a summer in Birmingham prior to beginning his university studies. He reported having lived in Wiltshire and Berkshire.

25. Previous research (e.g., Ganong, 1980) suggests that listeners would tend to identify a vowel token that was ambiguous between /e/ and /æ/ as hat rather than het because, of these two choices, only hat is a lexical item of English. Smaller response biases might result from differences in familiarity among real-word responses.
26. Data for two Dutch talkers were excluded to ensure an equal number of talkers in all three subgroups.

27. Another possibility is that both global pronunciation and specific aspects of segmental articulation depend on some general speech learning ability rather than amount of L2 experience. However, this seems unlikely since all 10 of the engineering students were assigned to the Strong Foreign Accent group, and all 10 of these talkers produced English /l/ with Dutch-like short-lag VOT values.

28. The same results were obtained when the values were transformed using an arcsine transformation recommended by Kirk (1968), and therefore will not be reported here.

29. The average rate for vowels spoken by native BE talkers ranged from 90% to 97% correct (listeners A2 and A1, respectively); and the rate for the Dutch talkers ranged from 85% to 88% (listeners A3 and A2, respectively).

30. The vowel effect was nonsignificant for the native BE talkers \[F(5,35) = 2.41, p = 0.056\]. As expected, they seem to have been equally successful in producing all six vowels. The effect of Vowel was significant, however, for the Dutch talkers with strong, moderate, and mild accents \[F(5,35) = 15.7, 4.16, 4.07; p < 0.01\]. Post-hoc tests showed that for the Mild Accent group, /u/ was less intelligible than /i/ and /l/. For the Moderate Accent group, /o/ but not /u/ was less intelligible than /i/ and /l/. For the Strong Accent group, both /u/ and /o/ were less intelligible than /i/ and /l/. In addition, /u/ and /o/ were less intelligible than /u/; and /æ/ was less intelligible than all other vowels.

31. The BE listeners' use of the keyword *hot* is marked as /o/ whereas the AE listeners' use of the same keyword is marked as /æ/.

32. For the AE listeners, the number of talkers showing the following variants for the Mild Accent group were: /æ/-3 talkers, /o/-2 talkers, /u/-1 talker; in the Strong Accent group it was: /æ/-8, /o/-3, /u/-1, /æ/-1). For the BE listeners the number of talkers in the Mild Accent group was /æ/-4, /æ/-3, /o/-1, /æ/-1, /u/-1, and in the Strong Accent group it was /æ/-8, /æ/-6, /o/-4, /æ/-1, /u/-1). There was not a significant difference in the rate at which the six vowels spoken by the native BE talkers were identified \[F(5,35) = 2.43, p > 0.10\], but a Vowel main effect was obtained for the Dutch talkers with mild, moderate and strong accents \[F(5,35) = 7.68, 4.24, 17.7, p < 0.01\]. For the mildly-accented talkers, /æ/ was identified less often than /i/ and /l/, their /æ/s were identified less often than /o/ or /æ/, and their /æ/ were identified less often than the /i/ or /l/. For the moderately-accented talkers, /æ/ was identified less often than /i/ and /l/. For the Dutch talkers with strong accents, both /æ/ and /æ/ were identified less well than /i/, /l/, /æ/ and /o/; and /o/ was identified less often than /i/ or /l/.

References


Elsendoorn, B. 1986. Personal communication.


L2 VOWEL INTELLIGIBILITY


——. 1989b. Differences in inventory size affects the location but not the precision of tongue positioning in vowel production, Language Speech, 32, 123-147.


James, A. 1985b. Personal communication.

James, A. 1989. Personal communication.

L2 VOWEL INTELLIGIBILITY


JAMES EMIL FLEGE


van Heuven, V. 1986. *Personal communication*.


JAMES EMIL FLEGE