Mimicry of Non-distinctive Phonetic Differences Between Language Varieties*

James Emil Flege  
University of Alabama, Birmingham

Robert M. Hammond  
Miami-Dade Community College

A delayed mimicry paradigm was used to assess speakers’ awareness of non-distinctive phonetic differences which in part distinguish languages. The notion of “phonological filtering” implies that second language learners may not be able to perceive phonetic differences between their native language and a foreign language unless the phonetic differences are linguistically relevant in the native language. If cross-language phonetic differences are in fact perceived poorly, it is unlikely that phonetic modification will occur in the course of naturalistic second language acquisition. In this study native English speakers familiar with Spanish-accented English attempted to read sentences with a Spanish accent. Acoustic measurements showed that two phonetic characteristics of English—the long VOT values associated with /p,t,k/ and final-syllable lengthening—were altered in the direction of Spanish and Spanish-accented English. These results provide tentative evidence that non-distinctive phonetic differences between languages are detectable by language learners and thus do not present an insuperable barrier to phonetic learning in second language acquisition.

For several decades phonologists have believed that “phonological filtering” contributes importantly to the presence of foreign accent in the speech of second language learners (Trubetzkoy 1969:51 ff). Having acquired the phonology of one language, the language learner will tend to interpret sounds heard in a foreign language in terms of phonological units found in the native language. For phonological filtering to occur phonological units in two different languages must be judged to be the “same” or “similar” (Weinreich 1953:7). A major problem for contact analyses in general and contrastive analyses, in particular, is to determine how such equivalencies are established by bilinguals, and at what level(s) of analysis the process of interlingual identification is best described.

Most investigations of second-language phonology have used either the allophone or the distinctive feature as the prime unit of analysis. Work done within a structural phonemic framework (e.g., Weinreich 1953; Shen 1959; Moulton 1962) has focused on between-language differences in phonemic inventory and their effect on the system of linguistically relevant contrasts, on the distribution of allophones, and on constraints affecting permissible sequences of surface phones. In the words of Robert Lado (1957:11), we tend to “transfer...our phonemes and their variants” when learning a foreign language. Between-language differences involving non-segmental characteristics of speech such as stress, intonation, and juncture have also been handled within a framework of phonemic contrast.
Other approaches have made use of the distinctive feature as the prime unit of analysis. Ritchie (1968) proposed that a perceptual hierarchy of distinctive features could be used to predict which native-language phone would be used in place of a foreign-language phone absent from the learner's native-language inventory (e.g., [t] or [s] for English /θ/). Eckman (1977) proposed that markedness, as established by cross-language implicational relations (typological markedness), might be used to predict how and where phonological contrasts will be neutralized by language learners. Still other investigations have focused on rules or processes relating surface phones to underlying phonological categories and their possible effect on pronunciation of foreign-language words (Lovins 1975; Eckman 1981).

An assumption that often seems to be made is that inter-lingual contact can be adequately described in terms of a small number of discrete linguistic categories: i.e., allophones, phonemes, or segmental changes describable in terms of distinctive features. Relatively little importance has been attached to the phonetic substance of surface phones found in the native and target languages in many theoretical accounts, except for its possible role in influencing which phonemes will be considered the "same" in the two languages. Once assigned a structural phonemic description, a target-language phone is expected to be replaced by a phonetically similar allophone of the corresponding native language phoneme or, if no such allophone exists, to be phonetically approximated. For example, learners may substitute the alveolar stop of the native language for a dental stop which occurs in a foreign language, or replace the uvular /t/ of a language they are learning with the approximate /t/ they learned as a child (Weinreich 1953).

It may be important, however, to pay somewhat closer attention to the phonetic characteristics of sounds in the native and foreign language (Brière 1966). Cross-language interference at the phonetic level probably contributes a great deal to what is perceived as foreign accent. Phonetic implementation of comparable phonological units in two languages may differ in important ways (Fromkin 1979; Ladefoged 1980). In such cases it appears that articulatory habits previously established to realize phonological units in the native language will be used in producing a foreign language (Flege 1981; Flege and Port 1981).

Phonological filtering may be at least partially responsible for cross-language phonetic interference, for knowledge of the native-language phonology may lead a second-language learner to ignore purely phonetic differences between languages. Bloomfield (1933:79), for example, proposed that speakers learn to attend to just those features of speech that are linguistically distinctive, ignoring other purely phonetic differences which are "redundant" or non-distinctive (see also Gibson and Levin 1975, for a more recent statement of this position). If this is true, language learners may fail to modify pre-existing phonetic habits when faced with cross-language differences that are not linguistically distinctive in their native language because they do not perceive such differences. For example, an American may make the phonemically long vowels of Swedish much shorter than do most Swedes—producing them with relatively shorter durations which approximate the durations of spectrally similar English vowels—because vowel duration is not "phonemic" (i.e., linguistically relevant) in English (Jonasson and McAllister
1972). On the other hand, learners may become aware of purely phonetic differences between their native language and a foreign language in the process of learning the foreign language. If this were the case, one might conclude that non-authentic pronunciation of foreign languages has primarily an articulatory rather than perceptual origin, resulting perhaps from general neurophysiological constraints on the acquisition or modification of fine motor skills rather than from perceptual limitations imposed by phonological filtering or previous linguistic experience (see Strange and Jenkins, 1978).

The present study was conducted in the hope of shedding some light on the ability of speakers to detect phonetic differences between languages. If it can be demonstrated that speakers in a language contact situation are aware of purely phonetic [i.e., "redundant" or "non-distinctive"] differences between language varieties it would suggest that phonological filtering does not impose an insuperable barrier to phonetic learning in second-language acquisition. Two phonetic differences between English and Spanish provided a means by which to assess the hypothesis that speakers are able to detect non-distinctive phonetic differences between language varieties.

**Spanish-English Phonetic Differences**

One way the voiceless stops of Spanish and English differ is according to VOT (voice onset time). English /p,t,k/ are produced with "long-lag" VOT values, whereas voicing begins soon after stop release for the unaspirated "short-lag" /p,t,k/ of Spanish (Lisker and Abramson 1964; Williams 1977). As a result, native English speakers often tend to produce voiceless Spanish stops with VOT values that are too long (i.e., with too much aspiration; Stockwell and Bowen 1965) while Spanish speakers predictably under-aspirate English /p,t,k/ (Williams 1979; cf. Jones 1948; Flege 1980).

There are several reasons to think that the kind of cross-language differences which lead to inappropriate VOT values during pronunciation of a foreign language will not be linguistically relevant [i.e., lead to the perception of segmental substitutions] for native-speakers experiencing a foreign accent. The distinctive difference between English /p,t,k/ vs. /b,d,g/ is ordinarily considered to be one of "voicing" or "tensity" rather than one of degree of aspiration or VOT. Although VOT may in fact be used to distinguish between natural classes of stops in English and other languages (e.g., /p,t,k/ vs /b,d,g/), VOT is not equivalent to the distinctive linguistic feature "voicing" or "tensity". In fact, degree of aspiration is ordinarily considered by phonologists to be a redundant property of English stops because it is predictable from phonetic context (see Port and Rotunno 1979; Weismer 1979a).

It is well known that manipulation of the VOT value of synthetic CV syllables (such as /pa/ vs /ba/) will have an important effect on whether the syllable is judged to be initiated by a voiced or voiceless stop (see, e.g. Abramson and Lisker 1970, 1973). However, other acoustic dimensions which co-vary with VOT in natural speech may also importantly affect stop voicing judgments made by English speaking listeners in the absence of clear VOT differences between stops (Lisker 1978a).
Thus it is not surprising that an English speaker might hear a short-lag (i.e. voiceless unaspirated) stop produced by a deaf speaker of English (Monsen 1976) or by an English-learning child (Eilers and Oller 1976, R. Eilers, personal communication) as phonologically voiceless (i.e., /p, t, k/) even though its VOT value might be less than the approximately 35 msec needed to cue voiceless stop judgments in perception experiments using synthetic speech stimuli. Similarly, phonetically naive English speakers are strongly inclined to label unaspirated Dutch or Korean stops as phonologically voiceless (i.e., as /p, t, k/), even though they are produced with VOT values which are much shorter than those typical of corresponding English voiceless stops (Lisker 1978b, cf. Brière 1966).

Even if VOT or degree of aspiration were a distinctive rather than redundant property of English stops, inappropriate VOT values in foreign-accented English might not lead to the perception of linguistically relevant differences. In most natural communicative settings context will serve to disambiguate an ambiguous phonetic contrast between minimal pairs like “to” vs “do”. At the same time, native English speakers might be expected to adapt their perceptual expectations when listening to Spanish speakers of English. It has been shown that Spanish-English bilinguals modify their judgments (as “voiced” or “voiceless”) of synthetic CV syllables differing in VOT as a function of whether they are listening in a Spanish or in an English mode (or “set”) to the stimuli (Elman, Diehl and Buchwald 1977, cf. Monsen 1978).

Another phonetic difference between Spanish and English pertains to the duration of syllables occurring in utterance-final position. In English the vocalic nucleus of utterance-final [pre-pausal] syllables may be 60-80% (about 100 msec) longer than the same vowel occurring in utterance-medial positions. Although this prosodic characteristic of English is acquired at an early age by children (Smith 1978) and has been found to be important for English speech to be judged as “natural” sounding (Carlson, Granström and Klatt 1979), it does not affect propositional meaning.

The magnitude of final-syllable lengthening is known to be considerably smaller in Spanish than English (Delattre 1966; Oller 1973, 1977). As a result of cross-language phonetic interference (Flege and Port 1981) native Spanish speakers seem to produce somewhat less final-syllable lengthening in English than do native speakers of English (Pinkerton 1973).

For the present study it is important to note that both the VOT and final-syllable lengthening differences between English and Spanish: (1) fall within the range of duration differences which are detectable by human listeners; (2) are considered “redundant” or “non-distinctive” phonetic differences in English (i.e., for monolingual English speakers); and (3) are more aptly described by scalar phonetic features than with binary distinctive features. If phonological filtering does not serve to obscure phonetic differences like these, both phonetic differences should be detectable by English speakers learning Spanish and by English speakers who are experiencing Spanish-accented English. In the present study a delayed-mimicry paradigm was used to assess awareness of these two phonetic differences by Americans familiar with Spanish-accented English.
**Methods**

The task used in the present study called for the delayed mimicry [i.e., from memory] of a variety of English rather than the direct and immediate imitation of some externally present model. College-aged subjects, all English native speakers, were instructed to read a typed list of 21 English sentences with what subjectively seemed to them to represent a "typical Spanish accent". That is, they were to produce English sentences as best they could like "someone whose native language is Spanish" without introducing unnatural pauses to convey the impression of accentedness. No demonstration of Spanish-accented English was given, nor were any explicit instructions provided concerning how one might produce the effect of accentedness.

We can be reasonably certain that the speakers in this study were familiar with Spanish-accented English since all were enrolled at the time of the study in a beginning-level Spanish language class taught by a native Spanish speaker who spoke English with a "Spanish accent". In addition, speakers' responses to a language-background questionnaire were used to further ensure that subjects had been exposed to Spanish-accented English. From the original population of 125, fifty speakers (25 males, 25 females) were randomly chosen who indicated that they:[1] were actually acquainted with native speakers of Spanish; and [2] had lived in Florida (home to many Spanish speakers) for at least five years.

The mimicked sentences were all of the form "The _____ is on the ______." The sentence-medial and -final blanks in the carrier phrase each contained one of the C(C)VC English words shown in Table 1 (e.g. "The book is on the wig."). The lexical items were chosen to include three tokens each of English sounds thought susceptible to mispronunciation by Spanish speakers of English. The apparent substitution of one English sound by another English sound may be the aspect of Spanish-accented English most evident to monolingual English speakers. It is known that estimates concerning the degree of Spanish accentedness made by native English speakers are correlated with the number of sound substitutions present in the speech sample (Brennan, Ryan, and Dawson 1975).

<table>
<thead>
<tr>
<th>Lexical Items</th>
<th>Substitute</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>nose, cheese, hose</td>
<td>s ɔ</td>
<td>141 [47%]</td>
</tr>
<tr>
<td>vice, veil, vase</td>
<td>b ɔ</td>
<td>129 [43%]</td>
</tr>
<tr>
<td>big, pig, wig</td>
<td>i l</td>
<td>127 [42%]</td>
</tr>
<tr>
<td>book, hook, crook</td>
<td>u ʌ</td>
<td>61 [20%]</td>
</tr>
<tr>
<td>shell, sheet, sheep</td>
<td>ɛ s</td>
<td>49 [16%]</td>
</tr>
<tr>
<td>bean, phone, bone</td>
<td>η ʌ</td>
<td>1 [0.3%]</td>
</tr>
<tr>
<td>tape, tube, toad</td>
<td>d t</td>
<td>0 [0.0%]</td>
</tr>
</tbody>
</table>

Table 1. Lexical items inserted into the carrier phrase "The _____ is on the ______." At right, the number of sound substitutions produced by 50 native English speakers reading the test sentences with a Spanish accent.
In addition to English sounds having predictable sound substitutes in Spanish-accented English, three words with pre-vocalic /t/ (tape, toad, and tube) were also included in the test sentences. These items provided a means of determining whether a non-distinctive phonetic difference between unaccented English and Spanish-accented English would lead to a segmental substitution ([d] for /t/), a phonetic modification ([t] for [tʰ]), or no modification at all by native English speakers mimicking Spanish-accented English. At the same time, the presence of each word in both sentence-medial and sentence-final position permitted the quantification of final-syllable lengthening in the speech of Americans mimicking a Spanish accent.

Speakers recorded the test sentences on Wollensak cassette tape recorders through fixed head-set microphones while seated in alternate booths in a language laboratory following a mid-term oral examination in Spanish. The record level of each machine was set before the experiment so that recordings for each speaker were of approximately equal intensity. Speakers paced themselves, and were permitted to repeat any sentence with which they were unsatisfied.

Two phonetically-trained listeners independently transcribed one sound in each of the key words found in the test sentences. One judge (JEF) is a native speaker of English who does not speak Spanish; the other (RMH) is bilingual in Spanish and English. Transcriptions by the two judges agreed in the great majority of cases (91.2%). The few cases (2.2%) where agreement could not be reached after further listening were submitted to a third judge for a final decision.

The frequency with which a substitute was heard for the English target sounds /zl/, /l/, /l/, /l/, /l/, /l/ and /l/ are presented in Table 1. Speakers produced a total of 508 segmental sound substitutions. Not included in Table 1 are the eleven cases where some other segmental substitution was heard.

The two judges heard no segmental substitution for /l/. This stop was transcribed as [lʰ] 70% of the time, and as a voiceless unaspirated stop ([t]) 30% of the time.

The frequency with which speakers produced segmental sound substitutions while mimicking a Spanish accent was used as the basis for selecting two sub-groups from the larger population for an instrumental phonetic analysis. It was hypothesized that speakers who demonstrate relatively great awareness of segmental properties of accented speech might also be most aware of non-segmental phonetic differences between language varieties. Accordingly, sentences of the ten speakers (Group A1) who produced the largest number of segmental substitutions (ranging from 17-27 per speaker) were considered separately from the sentences produced by the ten speakers (Group A2) producing the fewest segmental substitutions (0-4 per speaker). A control group (Group U) of ten native English speakers (5 males, 5 females) producing unaccented English was also established. These speakers, similar in background and age to speakers in Groups A1 and A2, read the same list of 21 sentences produced by speakers mimicking Spanish accent. The control group was recorded (Ampex, Model 602) in a sound-treated booth; a constant mouth-to-microphone distance was insured by means of a cephalostat.
The original recordings of speakers in all three groups were dubbed onto a Crown (Model 700) tape recorder for spectrographic analysis (Voiceprint, Model 700). Any variation in the strength of the original signal was largely neutralized during the dubbing process. Only the six sentences with tape, tube, and toad occurring in the sentence-medial or sentence-final blank were analyzed.

Three acoustic intervals were measured by hand to the nearest five msec from the spectrograms: (1) Voice-onset time (VOT) of /t/ was measured from the beginning of the transient noise burst signalling stop release, to the first vertical striation signalling onset of phonation; (2) ‘vowel’ duration of /owl/, /uw/ and /ey/ was measured from the onset of phonation following release of /t/ to the offset of energy in the region of F2 and F3 signalling implosion of a following stop; and (3) ‘sentence’ duration was measured from the beginning of the vowel in the sentence-medial test word to the end of the vowel in the sentence-final test word. This provided six measures each of ‘sentence’ duration, VOT, and ‘vowel’ duration for each speaker.

Results
The two groups of speakers mimicking Spanish accent (A1, A2) differed from speakers producing unaccented English (Group U) on all three phonetic dimensions that were examined. The mean duration of phonetic intervals (VOT, ‘vowel’ and ‘sentence’) produced by the three speaker groups are presented in Table 2.

<table>
<thead>
<tr>
<th>Speaker Group</th>
<th>Position</th>
<th>A1</th>
<th>A2</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Medial</td>
<td>44(15)</td>
<td>54(20)</td>
<td>81(16)</td>
</tr>
<tr>
<td>VOT</td>
<td>Final</td>
<td>46(21)</td>
<td>63(25)</td>
<td>89(14)</td>
</tr>
<tr>
<td></td>
<td>Medial</td>
<td>194(44)</td>
<td>202(55)</td>
<td>141(24)</td>
</tr>
<tr>
<td>VOWEL</td>
<td>Final</td>
<td>184(42)</td>
<td>209(48)</td>
<td>217(35)</td>
</tr>
<tr>
<td>SENTENCE</td>
<td></td>
<td>1208(180)</td>
<td>1182(270)</td>
<td>999(104)</td>
</tr>
</tbody>
</table>

Table 2. Mean duration in msec of phonetic intervals in the speech of three speaker groups. Each mean is based on the averaged values for ten speakers. Standard deviations are in parentheses.

First, the effect of speaker group on VOT was significant \( F(2, 27) = 14.09, p = 0.0006 \). Figure 1 displays the VOT values measured for each of the stops produced by speakers in the three groups. Note that VOT values in the stops produced by speakers in Group A1 and Group A2 ranged from 10 msec to over 110 msec, and that no stop produced by speakers of unaccented English (Group U) had a VOT value of
less than 50 msec. In fact, the most frequent VOT value for stops produced by speakers in both Group A1 and Group A2 (30-35 msec) was less than that for any stop produced by speakers in Group U.

The mean VOT of speakers producing unaccented English (85 msec) agrees closely with values reported previously for English stops in a similar phonetic context (e.g., Summerfield 1975; Port and Rotunno 1979). However, as can be seen in Figure 2, speakers mimicking a Spanish accent (Groups A1, A2) produced /t/ with VOT values averaging about 30 msec shorter than those produced by speakers in Group U.

Post-hoc tests revealed that the VOT values of both sentence-medial and sentence-final stops produced by speakers in Group U were greater than those produced by speakers in Group A1 (Tukey's HSD, alpha = .01). The speakers who produced fewer segmental substitutions when mimicking Spanish accent (Group A2) did not differ from Group U, nor was there a difference between the two Spanish accent groups (A1, A2).

The three speaker groups were alike in producing stops with slightly longer VOT values in sentence-final than in sentence-medial words \[F(1, 27) = 4.72, p = .038\].
The size of this effect—about 6 msec or 11%—is similar to that noted in a previous study by Summerfield (1975).

Second, Figure 3 reveals that the three speaker groups differed in terms of final-syllable lengthening. Group U made vowels an average 76 msec (54%) longer in sentence-final than sentence-medial words. Speakers in Group A2, on the other hand, made final vowels only 7 msec (4%) longer in sentence-final words; and vowels produced in sentence-final words by the other group of speakers mimicking Spanish accent (A1) actually averaged 10 msec (5%) shorter than sentence-medial vowels.

The lack of final-syllable lengthening in sentences produced by the speakers mimicking Spanish accent appears to be the basis for the significant interaction of the factor Speaker Group (A1, A2, or U) with the factor Position-in-Sentence (medial vs final) on the mean duration of vowels ($F(2, 27) = 14.04, p = .00007$). Post-hoc tests (Tukey's HSD, $\alpha = .01$) revealed that only the speakers producing unaccented English (Group U) made sentence-final vowels longer than sentence-medial vowels. This seems to be due primarily to the fact that the sentence-medial vowels produced by speakers mimicking Spanish accent (Groups A1, and A2) were significantly longer than those of Group U ($p < .01$). There were no between-group differences for sentence-final vowels.
Figure 3. Mean duration in msec of vowels found in sentence-medial and sentence-final words produced by three speaker groups. Each bar is based on 30 measurements.

Finally, inspection of Figure 4 reveals a difference between groups in the mean duration of sentences \( F[2, 27] = 3.34, p = .0502 \). Sentences produced by Group A1 and Group A2 averaged about 200 msec (20%) longer than sentences produced by speakers in Group U. Post-hoc tests revealed that Group A1 but not Group A2 produced significantly longer sentences than the unaccented group [U]. This difference may be due in part to the longer duration of utterance-medial vowels produced by speakers mimicking Spanish accent. It may also result from the presence of pauses. Twelve sentences produced by speakers in Group A1 had to be excluded from analysis because the presence of pauses caused them to exceed the length of a spectrogram (2.5 sec). However, a finer differentiation of sentences according to the presence or absence of pauses was not attempted because of the difficulty of distinguishing fluent from non-fluent pauses.

Discussion
The present study demonstrates the ability of speakers to modify non-segmental phonetic dimensions of their speech when mimicking a variety of English different from their own native dialect. Spanish speakers of English may produce English
stops with VOT values which are shorter than those typically seen in stops produced by native English speakers (Williams 1979). Spanish speakers of English are also known to produce a lesser magnitude of final-syllable lengthening than do native speakers of English (Pinkerton 1973). Both characteristics of Spanish-accented English appear to result from the carry-over from Spanish into English of Spanish-specific phonetic patterns (Flege 1980; Flege and Port 1981; cf. Jones 1948).

Figure 4. Mean duration in msec of sentences produced by three speaker groups. Number of sentences analyzed was 60 for Groups U and A2, 48 for Group A1.

The modification of both VOT and final-syllable lengthening by English native speakers asked to mimic Spanish-accented English suggests that speakers may be aware of non-distinctive phonetic differences distinguishing their own native dialect from other dialects or language varieties. It seems unlikely that the speakers of this study could have produced the kind of phonetic modifications seen here without some form of perceptual "awareness".

However, several objections might be made against interpreting the data of this study to mean that speakers are able to mimic purely phonetic differences between language varieties. First, speaking rate is known to influence VOT and vowel duration, so that the effects observed in this study may be an artifact of between-group differences in speaking rate. However, rate differences probably cannot account for the observed tendency of speakers mimicking Spanish-accented Eng-
lish to produce /t/ with VOT values that are shorter than usual for English. As speaking rate decreases and phonetic intervals increase, VOT values tend also to increase (Summerfield 1975; Port 1976). Speakers mimicking Spanish accent made sentences about 20% longer than speakers producing unaccented English. If rate were responsible for between-group differences in VOT we should have observed longer VOT values for Group A1 and Group A2 than for Group U. But what was actually observed was a substantial decrease in VOT by speakers mimicking Spanish-accented English. Nor is it likely that a more localized change in rate—that is, a difference in the duration of vowels immediately following the VOT intervals measured—can explain the between-group VOT differences. VOT values tend to be relatively long before vowels lengthened by phonetic factors (Port and Rotunno 1979; Weismer 1979a). The utterance-medial vowels produced by speakers mimicking Spanish-accented English in this study were significantly longer than those of speakers producing unaccented English, yet the VOT values of stops directly preceding these vowels were significantly shorter.

Second, one might argue that the absence of final-syllable lengthening in the English of native English speakers mimicking a Spanish accent may not truly be due to a suppression of this prosodic characteristic of English. The absence of final lengthening in sentences produced by speakers in Group A1 and Group A2 derives primarily from the relatively great length of their sentence-medial vowels rather than to the shortening of their utterance-final vowels (see Figure 3). Perhaps speakers prolonged sentence-medial vowels while planning the segmental sound substitution they would produce later in the sentence as part of their effort to effect a Spanish accent. This, rather than an intent to alter the ratio of sentence-final and sentence-medial vowels, might explain the absence of final lengthening by speakers mimicking Spanish accent.

Another possibility is that speakers in Group A1 and Group A2 sometimes introduced short pauses after sentence-medial test words, thereby placing medial vowels in pre-pausal position. Pauses following sentence-medial vowels were, in fact, perceptually evident in a number of cases. Accordingly, spectrograms were re-examined to determine if there was temporal acoustic evidence for pauses following sentence-medial test words. In about one-third (22/60) of the sentences there was a delay of more than 200 msec between the end of the vowel of the medial test word and onset of the following vowel [in "is"]). This included cases where speakers added a vowel to the test word [e.g., [tep]"it" "tape", [toda] "toad"]. Using 200 msec as a conservative criterion for the presence of pauses, data for the eight speakers who provided no perceptual or acoustic evidence of intra-sentential pauses (4 from A1, 4 from A2) were re-analyzed. Their sentence-final vowels averaged 190 msec (SD = 59 msec); and their sentence-medial vowels averaged 175 msec (SD = 64 msec). Although this represents somewhat more final lengthening than seen for either Group A1 or Group A2 taken as a whole, it is considerably less than the 76 msec (54%) of final lengthening produced by speakers of unaccented English (Group U). Moreover, it does approximate the magnitude of final lengthening found in languages other than English such as Spanish (Delattre 1966; Pinkerton 1973;
Thus the presence of intra-sentential pauses probably cannot explain the lack of final lengthening in sentences produced by speakers mimicking Spanish accent.

Finally, it is possible that speakers in this study adopted a general "foreign accent" mode instead of actually mimicking Spanish-accented English. Producing speech in a general "foreign accent" mode might consist in part of reducing the physiological complexity of speech sound production. For example, it has been proposed that long-lag (i.e., voiceless aspirated) stops are neurologically more complex than are short-lag (i.e., unaspirated) stops such as the /p,t,k/ of Spanish (Kewley-Port and Preston 1974). It also appears that the magnitude of final-syllable lengthening is considerably greater in English than in other languages (Delattre 1966; Oller 1977) so that this phonetic dimension might also be considered relatively "marked" or "difficult" for speakers. Thus it would be of great interest to determine if, for example, Cubans asked to mimic English-accented Spanish would produce more final-syllable lengthening than is characteristic of Spanish or increase VOT in pre-vocalic stops. Either finding would provide important additional support for the hypothesis that speakers do detect purely phonetic differences between language varieties such as those which may in fact distinguish foreign-accented speech from that produced by native speakers.

Data from this study are also relevant to a question raised by phoneticians concerning the nature of a speaker's control over VOT in speech production. Stops are generally regarded as falling into one of several modal ranges of VOT (the "lead", "short-lag," and "long-lag" categories). In running speech, however, the stops produced by individual speakers are often intermediate in value to idealized modal category values. It may be the case that the laryngeal devoicing gesture which results in the aspiration associated with pre-vocalic stops is "pre-programmed" (Weismer 1979b), and thus not under the conscious or voluntary control of speakers. Any overlap between VOT categories for stops observed in running speech could be considered to be the indirect and perhaps unintentional consequence of, for example variation in phonetic context, stress, or speaking rate. But if, on the other hand, speakers are able to exert control over the time course of the laryngeal devoicing gesture and thus the VOT value which is ultimately measured (Lisker and Abramson 1971:778) speakers should also be able to vary VOT in the absence of contextual variations.

The present data show that some speakers, at least, can produce stops with VOT values spanning the short-lag and long-lag ranges. This supports the observation that there is no physical limitation on VOT values (Lisker and Abramson 1971). It is interesting to note that speakers did not appear to be substituting one pre-established timing pattern for another when mimicking Spanish accent. The VOT values measured for their /t/, although far shorter than is typical for English, were almost always longer than is typical for English /d/. 
Conclusions
This study provides support for the hypothesis that speakers are tacitly aware of purely phonetic differences distinguishing their native dialect from other language varieties. Native English speakers who mimicked Spanish-accented English in the absence of an explicit external model either reduced or eliminated final-syllabic lengthening, a prosodic characteristic which is much less prominent in Spanish and Spanish-accented English than it is in English. The speakers mimicking Spanish accent also produced /ɪ/ with VOT values that were considerably shorter than is typical of English. They thereby approximated the short-lag stops of Spanish and the frequently underaspirated voiceless stops that are typical of Spanish-accented English.

Further research using delayed mimicry is needed to determine if an alternative interpretation, namely that speakers in this study were producing speech in a general “foreign accent” mode, might not better explain the present results. However there is reason to believe that the speakers in the present study actually did mimic phonetic characteristics of Spanish-accented English. First, both phonetic dimensions examined here were modified in the expected direction, (that is, they approximated Spanish phonetic patterns). Second, in the segmental analysis preceding this study, speakers produced many segmental substitutions associated with Spanish-accented English but very few that were not. Those who demonstrated the greatest awareness of Spanish accent (as determined by the number of segmental substitutions produced) also modified their speech most at the phonetic level.

Thus the present results suggest a tacit knowledge of phonetic differences between language varieties is part of the phonological competence of speakers familiar with multiple varieties of language. Speakers in this study seem to have been able to detect two of the non-distinctive phonetic differences which distinguish Spanish from English. From this we tentatively conclude that phonological filtering does not make it impossible for second language learners to achieve a phonetically accurate pronunciation of foreign languages.

Footnotes
* The authors would like to thank personnel of the Language Laboratory at the University of Florida for their cooperation and Alan Burns, Gerald Canter, and John Locke for comments on an early draft of this article. This research was funded in part by NIH grant NS 07107 to the Institute for the Advanced Study of the Communication Processes and by NIH grant NS 07100 to the Department of Communicative Disorders, Northwestern University.

Notes
1. Phonological filtering has much in common with the notion “categorical perception” discussed in speech perception research. Speech stimuli are said to be “categorically” perceived when a listener’s ability to discriminate between any pair of speech sounds is predictable on the basis of how each has been labelled (identified) by the listener. Truly “categorical” perception predicts, for example, that two stimuli which differ in VOT but are nonetheless both consistently labelled as /ɪ/ will not be reliably discriminated,
Mimicry of non-distinctive phonetic differences / Flege & Hammond

whereas another pair of stimuli differing by the same amount of VOT will be reliably discriminated from one another if one is often labelled /l/ and the other /r/. Speech perception research has shown, however, that within-category discrimination is generally better than predicted by category labelling (see Liberman, Harris, Hoffman, and Griffith 1957) and can even be improved with training [Carney, Widen, and Viemeister 1977]. This is intuitively plausible. If such were not the case it would be impossible, for example, to detect a “distorted” production of /l/ or to detect some cross-dialect differences.

2. The possible substitution of one sound by another in Spanish-accented English is to some extent understandable through a comparison of Spanish and English. In Spanish there are no lax vowels such as English /u/ or /u/; orthographic “v” represents either [b] or [b], depending on phonological environment; and [z] does not occur either intervocally or in utterance-final position. Further, in Caribbean dialects of Spanish [c] and [s] may vary as surface realizations of /l/, and final /n/ may be phonetically realized as [ŋ].

References


Lisker, L. 1971. Distinctive features and laryngeal control. Language 47.767-785.


Shen, Y. 1959. Some allophones can be important. Language Learning 9.7-18.


