The purpose of this study, which focused on the word *taco* as spoken in Spanish and English, was to explore the word as a unit in second language (L2) speech acquisition. As expected, acoustic measurements revealed that Spanish and English monolinguals' renditions of *taco* differed systematically. It was also shown that the extent to which Spanish/English bilinguals approximated English phonetic norms for any one segment of *taco* was correlated with their approximation for the other three segments, and that early learners differentiated Spanish versus English *taco* more than did late learners. It also appeared that the bilinguals produced /t/ with less English-like voice onset time (VOT) values in English *taco* than in other English words without a cognate in Spanish. In a perception experiment, listeners were able to identify the native language of Spanish and English monolinguals on the basis of their production of *taco*. The listeners heard larger differences between Spanish and English *taco* tokens spoken by early than late learners of English L2. Two additional perception experiments assessed further the phonetic dimensions that listeners use to determine language identity and to gauge bilinguals' speech production accuracy. Listeners assigned to language identification and goodness rating tasks responded to acoustic information distributed over all four segments in *taco*, although the VOT of the word-initial /t/ appeared to be the single most important phonetic dimension. Taken together, the results of this study suggest that (a) bilinguals' accuracy in producing the various segments of a second language word may be interrelated and (b) in judging L2 speech, listeners respond to phonetic errors distributed over the entire word.
Most research examining the pronunciation of a second language (L2) falls into one of three broad categories. Some research has examined segmental production, either through phonetic transcriptions (e.g., Flege & Davidian, 1984), the identification of sounds in terms of native-language categories by nonphoneticians (e.g., Cochrane, 1980; Gatbonton, 1983), or acoustic measurements (e.g., Flege, 1991; Major, 1987, 1992). Other research has focused on prosodic aspects of L2 speech production, such as rhythm and/or intonation (e.g., Wenk, 1985). Still other research has examined overall degree of foreign accent in sentence- or paragraph-length materials (e.g., Flege & Fletcher, 1992; Oyama, 1976; Tahta, Wood, & Lowenthal, 1981). It is generally agreed that divergences from L2 phonetic norms for both segmental articulation and prosody contribute to foreign accent.

Many psycholinguistic studies have focused on the word as a unit of analysis, raising issues such as whether bilinguals possess single or dual lexicons (e.g., Kolers, 1965, 1966; Macnamara & Kushnir, 1971; Soares & Grosjean, 1984). Theorists have pointed to the importance of the sound segment (e.g., Flege, 1991) and syllable (e.g., Briere, Campbell, & Soemarmo, 1983) as units of control in L2 speech production. However, apart from studies examining the cognate status of individual L2 lexical items, few studies have examined the production of whole words in a second language. Thus, it is not known if L2 learners can “substitute” a single sound in a second language word independently of their production of neighboring sounds in the same word. The two primary aims of the present study were therefore to (a) examine the role of the word in L2 speech production and (b) assess listeners’ ability to perceive nonnative phonetic characteristics in the multiple segments within a word.

The lack of attention to the word in L2 production studies is surprising in view of the prominence given to this unit of analysis in studies of native language (L1) acquisition. It is generally agreed that the word is the child’s “earliest phonological unit” and that children initially tend to treat each new word as a “phonological whole” rather than as a “sum of segmental parts” (Menn, 1981, p. 135). Research in child L1 acquisition has shown that a child’s production of a particular sound may vary from word to word and that phonological processes may not affect a particular sound equally in all words (Ferguson & Farwell, 1975; Macken & Ferguson, 1981). Even beyond the time a child has acquired the metaphonological ability to segment words into individual vowels and consonants, the lexicon may continue to play an important role in segmental articulation. For example, young monolingual English children may produce /s/ more accurately in high- than low-frequency words (Leonard & Ritterman, 1971).

Faced with the need to understand what is being said to them, adult learners of a second language often seek—and sometimes find—lexical correspondences between words in the L2 and words found in the L1. These “old friends” arise through borrowings (either from the L1 into the L2, or vice versa) or from mutual inheritance from a common language ancestor. The term cognate will be used here to refer to any word in the L2 that is judged by the L2 learner—whether correctly or incorrectly—to be the “same” as a word in the L1. Although the phonological match between words in two languages is seldom if ever perfect, correspondences noted between L1 and L2 words are more likely to involve the sound substance of lexical
items rather than meaning or etymological history (Carroll, 1992). Psycholinguistic studies have shown differences in the time taken to recognize cognate and noncognate words (Caramazza & Brones, 1979), as well as priming effects on cognate words in a lexical decision task (Cristoffanini, Kirsner, & Milech, 1986). Also, L2 speech production research has suggested that phonetic segments may be produced less accurately in cognate than noncognate words (Cochrane, 1980; Hammerly, 1982) and that rhythm may also be less accurate in cognate than noncognate words (Wenk, 1985).

The effect of cognate status was tested in the present study by examining the production of word-initial /t/ by native speakers of Spanish and English. Flege (1991) found that native Spanish speakers who began learning English as adults produced /t/ in English words with significantly shorter voice onset time (VOT) values than did native speakers of English. (Their VOT values were therefore Spanish-like.) On the other hand, native Spanish speakers who had begun learning English by the age of 6 years (called “early learners”) did not differ from native speakers in producing English /t/. No attempt was made to distinguish stop consonants produced in cognate and noncognate words, so data from the Flege (1991) study were reexamined here. We hypothesized that the Spanish/English bilinguals would produce English /t/ with shorter—and thus Spanish-like—VOT values in cognate than noncognate words.

All five experiments of the present study focused on the word taco, which occurs in both Spanish and English. The Spanish and English renditions of taco might be regarded as consisting of the “same” four phonemes realized in phonetically different ways (e.g., Dalbor, 1980; Delattre, 1966; Flege, 1989; Flege & Eefting, 1986). Spanish /t/, for example, is realized with a dental place of constriction and short-lag VOT values, whereas English /t/ is realized with an alveolar place of constriction and long-lag VOT. The aims of two speech production experiments were to determine if Spanish/English bilinguals produce /t/ less accurately in cognate than noncognate English words; to determine what relationship, if any, exists in the accuracy with which the four segments of English taco are produced; and to learn if early learners produce a more robust distinction between their Spanish and English renditions of taco than do late learners. Results of the production experiments raised a number of questions that were addressed in three perception experiments: Can listeners identify the language in which taco has been spoken solely on the basis of the phonetic differences that distinguish the Spanish and English renditions of taco? If so, would they be better able to differentiate the Spanish and English renditions of taco that were spoken by early rather than late learners? And, do listeners make use of acoustic information in all four phonetic segments of taco when making cross-language phonetic judgments?

**EXPERIMENT 1**

Acoustic measurements were made of the word taco as spoken in either Spanish or English by monolinguals, and in both languages by Spanish/English bilinguals differing in age of L2 learning.
Method

Talkers. The subjects were a subset of those examined by Flege (1991). Each of the four groups of subjects in the present experiment was comprised of four adult males and three adult females. The mean ages of the Spanish and English monolinguals were 33 and 27 years, respectively. The English monolinguals were students at the University of Texas; the Spanish monolinguals had recently arrived in Austin, Texas. The “early learners” were native speakers of Spanish who began learning English in school at the age of 5 or 6 years, whereas the “late learners” did not begin learning English until adulthood. The early and late learners had arrived in the United States at average ages of 2 and 19 years, respectively. Compared to the early learners, the late learners had lived for a shorter period of time in the United States (13 vs. 20 years), were somewhat older (32 vs. 22 years), had received less formal education in English (5 vs. 12 years), and reported speaking English somewhat less often on a daily basis (60 vs. 80%).

Procedure. As described by Flege (1991), the monolinguals produced *taco* only in their native language whereas the bilinguals produced *taco* in both of their languages, in counterbalanced order. The subjects were recorded in a sound booth using a cassette tape recorder (Marantz PMD 420). Instructions were administered in Spanish or English to the monolingual subjects by a bilingual research assistant, and in either Spanish or English to the bilingual subjects. The subjects were asked to speak naturally at a constant, normal speaking rate. Production was elicited using written lists of Spanish and English words, both containing 27 disyllabic words. The only word contained on both the Spanish and English lists was *taco*. Just the second of the three available tokens were examined here.

Acoustic Measurements. Of the 42 *taco* tokens that were analyzed acoustically, 7 were spoken in English by English monolinguals, 7 in Spanish by Spanish monolinguals, 14 in Spanish by Spanish/English bilinguals (half early learners, half late learners), and the remaining 14 in English by the same bilinguals. Utterances containing the selected tokens were filtered (60–4000 Hz), digitized (10.0 kHz, 12-bit resolution), and stored on disk. The *taco* tokens were edited out of the Spanish or English carrier phrase in which they had been spoken (*Tengo un _, Take a _*), then normalized for peak intensity. The normalization procedure neutralized intensity differences across tokens in the first (stressed) vowel but preserved the relative intensity differences between the first and second vowels (which was always less intense than the first vowel).

A waveform editor was used to measure VOT from the beginning of the release burst of the stop consonants /t/ and /k/ to the first positive peak in the following vowels. The duration of the first vowel was measured from the first positive peak in the periodic portion of the waveform to constriction of the medial /k/. Duration of the second vowel was measured from the first positive peak to the last positive peak in the waveform. The closure duration of the intervocalic /k/ was measured from constriction onset to offset. A determination of the onset of constriction was based
on an abrupt decrease in the complexity of the waveform and a marked decrease in amplitude. Release of constriction was based on the occurrence of a release burst. The amplitude of Vowel 2 relative to the more intense Vowel 1, and the amplitude of the /k/ release burst relative to the peak intensity of Vowel 2, were measured in decibels (dB). Vowel formant frequency was estimated using linear predictive coding (LPC) analysis. The number of LPC coefficients computed was usually 14 for males and 12 for females. The frequencies of the first three formants (F1–F3) in Vowel 1 were computed by placing a 25.6-ms Hamming window at the acoustic midpoint. Measurements were made at just one location because preliminary analyses had revealed little evidence of formant movement. Formant frequencies were measured at the 20 and 80% point in all tokens of Vowel 2, which showed more evidence of formant movement than Vowel 1 (at least as spoken by some native English speakers).

Results and Discussion

**Analysis of Phonetic Segments.** The mean temporal and intensity values of Spanish and English renditions of *taco* are presented in Table 1. Mean frequency values are presented in Table 2. A series of one-way ANOVAs was carried out to test for differences between the two monolingual groups and the bilinguals' Spanish and English productions (i.e., the Group factor had six levels). The results of these analyses, as well as the results of Newman-Keuls post-hoc tests, are also presented in Tables 1 and 2.

As expected, the English monolinguals produced /t/ with significantly longer VOT values than the Spanish monolinguals (58 vs. 18 ms; *p* < .05). The early learners approximated English temporal phonetic norms to a greater extent than the late learners. The early and late L2 learners produced /t/ with longer VOT in English than Spanish *taco* (early learners: 36 vs. 16 ms; late learners: 20 vs. 13 ms), but only the early learners' difference was significant (*p* < .05). The early learners produced /t/ in English *taco* with significantly longer—and thus more accurate—VOT values than did the late learners (*p* < .05). No between-group difference reached significance for the VOT of /k/, but the values for this variable will nevertheless be of interest (see below). A significant Group effect was observed for the closure duration of /k/. The English monolinguals' /k/ was significantly shorter than that of the Spanish monolinguals (73 vs. 108 ms). Both groups of bilinguals produced /k/ with somewhat shorter closure in English than Spanish. The early learners produced /k/ in English *taco* with a significantly shorter (and thus more accurate) /k/ closure duration than did the late learners.

The subjects in both monolingual groups produced greater stress on the first than second syllable. Few between-group differences in intensity were observed. The effect of the Group factor on /k/ burst amplitude was nonsignificant, but the effect of Group on Vowel 2 amplitude was significant. The English monolinguals produced Vowel 2 with significantly greater intensity than the Spanish monolinguals (−0.9 vs. −5.3 dB). The late learners produced Vowel 2 with the same relative amplitude in
Table 1. Mean value of acoustic variables in segments of Spanish or English renditions of the word *taco*

<table>
<thead>
<tr>
<th></th>
<th>English <em>taco</em></th>
<th>Spanish <em>taco</em></th>
<th>Prob.</th>
<th>Post-hoc</th>
</tr>
</thead>
</table>
|                          | 1—Mono^
| Vot of /t/              | 58             | 18    | .000    | 1 > 2, 3, 4, 5, 6 |
|                          | (12)           | (8)            |       |          | 2 > 3, 4, 5, 6   |
| Vot of /k/               | 36             | 45             | .592  |          |
|                          | (13)           | (12)           |       |          |
| Vowel 1 duration         | 118            | 113            | .900  |          |
|                          | (18)           | (21)           |       |          |
| Vowel 2 duration         | 149            | 63             | .002  | 1 > 3, 4, 6 |
|                          | (24)           | (22)           |       |          |
| V1/V2 duration ratio     | 0.80           | 1.98           | .073  |          |
|                          | (0.2)          | (0.7)          |       |          |
| /k/ closure duration     | 73             | 108            | .000  | 4, 6 > 1, 2, 5 |
|                          | (27)           | (21)           |       | 3 > 2, 5  |
| Vowel 2 intensity        | -0.9           | -5.3           | .002  | 4 > 1, 2, 5 |
|                          | (0.6)          | (2.8)          |       |          |
| Intensity of /k/ burst   | -6.8           | -6.7           | .991  |          |
|                          | (6.1)          | (5.0)          |       |          |

Note: Duration measures are in milliseconds; intensity measures in dB. Standard deviations are in parentheses. Prob. indicates the probability that the effect of Group was significant in a one-way ANOVA; Post-hoc indicates between-group differences significant at the .05 level according to a Newman-Keuls test.

Table 2. The mean frequencies of the first three formants (F1-F3) in the two vowels of taco as spoken in English or Spanish

<table>
<thead>
<tr>
<th></th>
<th>English taco</th>
<th></th>
<th>Spanish taco</th>
<th></th>
<th></th>
<th>Prob.</th>
<th>Post-hoc</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-Mono</td>
<td>2-Early</td>
<td>3-Late</td>
<td>4-Mono</td>
<td>5-Early</td>
<td>6-Late</td>
<td></td>
</tr>
<tr>
<td>F1 of Vowel 1</td>
<td>722 (151)</td>
<td>782 (120)</td>
<td>745 (109)</td>
<td>710 (63)</td>
<td>772 (121)</td>
<td>748 (71)</td>
<td>.819</td>
</tr>
<tr>
<td>F2 of Vowel 1</td>
<td>1324 (176)</td>
<td>1364 (198)</td>
<td>1437 (157)</td>
<td>1476 (128)</td>
<td>1420 (175)</td>
<td>1505 (125)</td>
<td>.313</td>
</tr>
<tr>
<td>F3 of Vowel 1</td>
<td>2310 (193)</td>
<td>2481 (132)</td>
<td>2634 (154)</td>
<td>2561 (251)</td>
<td>2552 (166)</td>
<td>2678 (159)</td>
<td>.014</td>
</tr>
<tr>
<td>F1 of V2 (onset)</td>
<td>599 (105)</td>
<td>520 (79)</td>
<td>487 (128)</td>
<td>518 (64)</td>
<td>453 (81)</td>
<td>525 (227)</td>
<td>.376</td>
</tr>
<tr>
<td>F2 of V2 (onset)</td>
<td>1483 (221)</td>
<td>1100 (203)</td>
<td>901 (112)</td>
<td>866 (77)</td>
<td>903 (88)</td>
<td>896 (167)</td>
<td>.000</td>
</tr>
<tr>
<td>F3 of V2 (onset)</td>
<td>2285 (214)</td>
<td>2420 (294)</td>
<td>2471 (262)</td>
<td>2500 (127)</td>
<td>2206 (289)</td>
<td>2339 (143)</td>
<td>.168</td>
</tr>
<tr>
<td>F1 of V2 (offset)</td>
<td>544 (114)</td>
<td>538 (143)</td>
<td>522 (250)</td>
<td>546 (124)</td>
<td>504 (58)</td>
<td>592 (242)</td>
<td>.981</td>
</tr>
<tr>
<td>F2 of V2 (offset)</td>
<td>1287 (216)</td>
<td>1044 (134)</td>
<td>874 (197)</td>
<td>841 (81)</td>
<td>886 (93)</td>
<td>900 (233)</td>
<td>.001</td>
</tr>
<tr>
<td>F3 of V2 (offset)</td>
<td>2319 (235)</td>
<td>2434 (215)</td>
<td>2542 (138)</td>
<td>2328 (132)</td>
<td>2572 (267)</td>
<td>2289 (250)</td>
<td>.120</td>
</tr>
</tbody>
</table>

Note: Frequencies are in Hz. Standard deviations are in parentheses. Post-hoc indicates between-group differences significant at the .05 level according to a Newman-Keuls test.

Spanish and English taco. The early learners produced Vowel 2 with nonsignificantly greater intensity in English than in Spanish. The English monolinguals produced the second vowel of taco with significantly longer duration than the Spanish monolinguals (149 vs. 63 ms). This does not belie the observation made earlier concerning stress, for the difference was probably the result of greater utterance-final lengthening by the English than Spanish monolinguals (Oller, 1979). Both bilingual groups produced longer Vowel 2 durations in English than Spanish (early learners: 120 vs. 103 ms; late learners: 90 vs. 73 ms), but neither cross-language difference was significant ($p > .05$). However, whereas the late learners differed significantly ($p > .05$) from the native English speakers in producing the second vowel of English taco, the early learners did not.

A significant effect of Group was observed for just two of the nine frequency variables. The effect of Group was significant for the onset frequency of F2 in Vowel 2. The English monolinguals' F2 was significantly higher than that of the Spanish monolinguals (1483 vs. 866 Hz). The late learners increased F2 frequency by a nonsignificant 5 Hz when switching from Spanish to English, and the early learners by a nonsignificant average of 197 Hz. A significant effect of Group on the offset frequencies of the F2 in Vowel 2 was obtained, however. Once again, frequencies were significantly higher for the English than Spanish monolinguals (1287 vs. 841 Hz). The late learners showed a nonsignificant decrease in F2 when switching from Spanish to English and thus diverged further from the phonetic norm of English. The early learners showed an appropriate, but nonsignificant, increase in F2 frequency when switching from Spanish to English.

In summary, the acoustic analyses revealed a number of significant differences in the production of taco by Spanish and English monolinguals. In many instances, bilinguals showed modifications in the appropriate direction when switching from Spanish to English. As expected, early learners produced larger differences between Spanish versus English taco than did the late learners. In only one instance (i.e., VOT in word-initial /t/) did the early learners produce a significant Spanish versus English difference. In just two instances (the VOT of /t/; the closure duration of /k/) were the early learners' productions significantly more English-like than those of the late learners. The paucity of significant differences between the early and late learners came as a surprise (Flege, 1991). As suggested in the introduction, one possible explanation for this is that segmental production of L2 sounds by bilinguals may be less accurate in L2 words that have a cognate in the L1 than in L2 words without a cognate in the L1. This possibility was tested in Experiment 2.

**Segment Interrelations.** A correlational technique was used to determine whether the Spanish/English bilinguals produced certain segments in English taco accurately and others inaccurately, or their accuracy in producing the four segments of taco varied as a function of the word as a whole. We examined the correlations of all 120 possible pairings of 16 acoustic variables measured in the 28 Spanish and English taco tokens that had been spoken by the bilinguals. Of these, 36 were significant at the .05 level. However, of the significant correlations, only 12 involved two different acoustic dimensions (e.g., duration, frequency) or two different seg-
ments and also involved an acoustic variable that differed significantly in the Spanish and English monolinguals' productions of *taco*.

The 12 correlations meeting the two criteria just mentioned are listed in Table 3. For all 12, as the value for one variable became more English-like (i.e., moved from the mean observed for Spanish monolinguals toward the mean observed for English monolinguals), the value for the other variable did so also. Take, for example, the relation between VOT in the word-initial /t/ and closure duration of the intervocalic /k/. The English monolinguals produced /t/ with significantly longer VOT values than did the Spanish monolinguals, but they produced the closure duration of /k/ with significantly shorter durations than did the Spanish monolinguals (see Table 1). For the bilinguals' productions of *taco*, a significant negative correlation existed between the VOT of /t/ and the closure duration of /k/, \( r(26) = -.419, p < .05 \). That is, the longer (more English-like) their production of VOT in /t/ was, the shorter (more English-like) were their productions of the closure duration of /k/. A similar pattern was evident for the 11 other significant correlations listed in Table 3. Notice that the list includes pairs of variables from adjacent segments (e.g., /k/ closure duration–Vowel 2 intensity) as well as from nonadjacent segments (e.g., /t/ VOT–F2 frequency of Vowel 2). In no instance did the value for one variable move away from the Spanish monolingual mean toward the English monolingual mean (what we call *phonetic approximation*) with the value for the other variable in a pair moving away from the monolingual English mean.

The pattern of correlations in Table 3 suggests that cross-language phonetic interference in the production of L2 words may affect *whole words*, not just individual sounds. This finding is more consistent with the general approach of nonlinear phonology than that of taxonomic phonemics. A nonlinear phonological approach regards phonetic segments as temporal "place holders" in larger production units, rather than as well-specified units that are explicitly targeted by the talker (e.g.,
Goldsmith, 1990). Alternatively, VOT in a stop consonant such as /t/, for example, might be regarded as being determined by global production parameters apt to influence multiple segments rather than by temporal implementation rules operating on a single, explicit segment (Port, Dalby, & O'Dell, 1987, p. 1584) or by some global change in articulatory setting (e.g., Esling & Wong, 1983).

Still another explanation of the intercorrelations in Table 3 was suggested by the results obtained in a switching experiment. Grosjean and Soares (1986) examined English words inserted into French sentences by French/English bilinguals. In an utterance such as On a brunché avec eux (“We had brunch with them”), the first part of the inserted English word (viz., brunch) revealed French acoustic phonetic characteristics, whereas later portions of this word had English acoustic phonetic characteristics. Bilinguals may produce a phonetic segment in a second language word in a target-like fashion by using newly acquired realization rules (Flege, 1991; Flege & Port, 1981; Port & Mitleb, 1983). In so doing, they may inadvertently alter the production of nearby segments in the second language word. This assumes that the time needed to switch between different language-specific realization rules is longer than the duration of an average phonetic segment. Some support for this is provided in Experiment 2.

**EXPERIMENT 2**

The bilinguals in Experiment 1, especially the early learners, produced the /t/ in English taco with shorter VOT values than was expected (Flege, 1991), perhaps because taco also occurs in Spanish. Hammerly (1982) reported that Spanish/English bilinguals produced more segmental errors in English words that had a cognate in Spanish than in noncognate English words (33 vs. 8%). Cochrane (1980) reported that Japanese adults produced English /a/ and /l/ less accurately in cognate than noncognate words. No previous study, to our knowledge, has used acoustic measurements to compare the accuracy of sounds in cognate versus noncognate words. The aim of this experiment was therefore to determine whether the Spanish/English bilinguals from Experiment 1 would produce English /t/ with shorter (i.e., more Spanishlike) VOT values in cognate than noncognate words. VOT was measured in their production of the word-initial /t/ in the English word taco, as well as in other English words expected to differ in terms of the perceived relationship to words in the Spanish lexicon.

**Method**

The methods used here have been described in detail by Flege (1991). In the earlier study, the early and late learners from Experiment 1 (along with six others) read an alternating list of disyllabic Spanish and English words beginning in /t/ in three speaking conditions. The design afforded an opportunity to assess how rapidly and accurately bilinguals could modify their production of /t/ when switching with increasing rapidity between Spanish and English. In the first condition, the words on the list were produced at the end of a Spanish or an English sentence (Take another
word such as _, Tengo palabras como _), as appropriate. In the second speaking condition, the words were produced at the end of an English or Spanish phrase (Take a _, Tengo un _). In the third and final speaking condition, the Spanish and English words were simply read in isolation. As a result of this manipulation, the interval between the word-initial /t/ in successive Spanish and English words varied. The Spanish and English utterances were highlighted in different colors so that language identity was never in question, even in the isolated word condition.

The list of English words included taco and other words in which word-initial /t/ was followed by /i/ (teabag, teeshirt, teapot, t-bone) or /e/ (temple, tempo, textbook, teller). The list of Spanish words also included taco and words in which the word-initial /t/ was followed by /i/ (tigre, tipo, tiro, timbre) or by an [e] or [e] realization of /e/ (tema, termo, templo, texto). We thought it unlikely that the bilinguals would relate any of the English words with /i/, or the word teller, to a word in Spanish. They were more likely, we thought, to relate English textbook to the Spanish word texto, and very likely to relate the English words temple and tempo to the Spanish words templo and tiempo, respectively. These assumptions were not verified at the time of the experiment, however.

Flege (1991) did not measure VOT in taco. In that study, a single mean VOT value was reported for the words in which /t/ was followed by /i/ or /e/. To the VOT measurements obtained in the earlier study, we added three measurements of the word temple for each talker (one for each speaking condition) and 12 measurements of taco (2 languages x 3 speaking conditions x 2 tokens). Nine mean VOT values were calculated for each subject's production of English /t/: one for taco, one for the English words with /i/, and one for the English words with /e/ in each of the three speaking conditions. Similarly, nine mean values were calculated for the /t/s in Spanish words.

Results and Discussion

The mean VOT values were submitted to a Group (early vs. late learners) x Language (Spanish vs. English) x Cognate Status (taco, words with /i/, words with /e,e/) x Speaking Condition (words in sentences, words in phrases, isolated words) ANOVA. The last three factors were repeated measures. If cognate status affects the accuracy with which sounds are produced in a second language, then a Language x Cognate Status interaction should be obtained. That is, one would expect the bilinguals to produce a larger Spanish versus English VOT difference for the words with /i/ than in taco. This is because none of the English words with /i/ has a Spanish cognate, whereas English taco clearly does. A significant two-way interaction was in fact obtained, $F(2, 24) = 32.0, p < .01$. The bilinguals produced /t/ with longer VOT values in the English /i/ words than in the English words with /e/ or in English taco (55 vs. 44 vs. 26 ms). A post-hoc test revealed that all three mean VOTs differed significantly ($p < .01$). The bilinguals produced /t/ with slightly longer VOT values in Spanish /i/ words than in either Spanish /e/ words or in Spanish taco (24 vs. 19 vs. 19 ms). The difference between Spanish /i/ words and Spanish taco was significant but not, of course, the difference between /e/ words and taco ($p > .05$).
The much larger Spanish versus English VOT difference for words with /i/ than for taco (31 vs. 7 ms) is consistent with the hypothesis that certain L2 sounds will be produced less accurately in cognate than noncognate words. However, caution is needed in interpreting this finding. This is because the bilingual subjects’ perception of the degree of relatedness between the English words examined and words in the Spanish lexicon was not assessed. Also, VOT is often somewhat longer in stops preceding high than nonhigh vowels, and cognate status was confounded here with vowel height (Lisker & Abramson, 1967; Port & Rotunno, 1979; Weismer, 1979). An aerodynamically based vowel height effect may thus have contributed to the “cognate” effect. It is unlikely to have been entirely responsible for it, however. Although the effect on VOT that arises from the tongue configuration for an upcoming vowel should be much the same in Spanish and English, the bilinguals produced a much larger VOT difference between English words with /i/ and /a/ (viz., English taco) than between the Spanish words with /i/ and /a/ (28 vs. 5 ms), t(13) = 5.92, p < .01. Still, we readily admit that additional research is needed to evaluate the role of cognate status on L2 segmental production accuracy.

The ANOVA also yielded a significant Group x Language interaction, F(1, 12) = 12.0, p < .01, because the early learners produced a much larger English versus Spanish VOT difference than did the late learners (early learners: 51 vs. 21 ms; late learners: 32 vs. 21 ms). However, even the late learners’ relatively small Spanish versus English VOT difference was significant (p < .05). Finally, the ANOVA yielded a significant Speaking Condition x Language interaction, F(2, 24) = 8.10, p < .01, because larger English versus Spanish differences were evident in the sentence and phrase conditions than in the isolated word condition (sentence: 45 vs. 20 ms; phrase: 40 vs. 19 ms; word: 40 vs. 24 ms). The difference between the three conditions may have arisen because the amount of time between successive Spanish and English /s/ decreased progressively across the three conditions, leaving less time for the bilinguals to switch between sets of realization rules appropriate for Spanish and English (see Flege, 1991; Flege & Eefting, 1987, 1988). Alternatively, the decreasing interval across the three conditions between successive Spanish and English /s/ may have provided less opportunity for the bilinguals to reset longer term production parameters.

EXPERIMENT 3

Experiment 1 identified a number of phonetic differences in the Spanish and English renditions of taco. It also revealed a number of effects of cross-language phonetic interference on Spanish/English bilinguals’ productions of taco in English. The present experiment and the two that follow examined listeners’ perception of the phonetic differences between Spanish and English monolinguals’ renditions of taco and between productions of English taco by native English and Spanish speakers.

One question of interest was whether listeners could determine language identity solely on the basis of the phonetic differences that distinguish the Spanish and English renditions of taco. Some models of speech perception such as Elman’s
TRACE model (Elman, 1989) posit separate phonetic and phonemic levels of perceptual processing. The language-specific phonetic level is regarded as more detailed and as occurring earlier in processing than the phonemic level. The extent to which phonetic-level information is retained after phonemic identification and word recognition has taken place may vary as a function of processing task. If listeners in the present experiment were unable to access phonetic-level information, they might find it difficult to identify the language in which a taco token had been uttered. Conceivably, listeners might need to know the language “identity” of a word in order to access and/or use phonetic-level information. Obler and Albert (1978) hypothesized that bilinguals monitor constantly the identity of the language being spoken and that language identity judgments are based on the detection of certain elements (phonemes, phoneme sequences, words, syntactic structures) that exist in one language but not the other. No such cues were available in the taco tokens presented to listeners in the present experiment.

Other previous research supports the view that phonetic details remain perceptible. For example, at least some of the phonetic differences distinguishing two dialects of a language can be acquired (Chambers, 1992), which implies that they are detected and stored in long-term memory. Flege (1984) showed that listeners can detect small divergences from L1 norms in foreign-accented speech that are localized in a single segment. Grosjean and Soares (1986) found that French/English bilinguals could distinguish a uvular French /R/ from the dorsal approximant English /J/. An increasingly larger portion taken from the beginning of a French syllable was replaced by corresponding portions from an English syllable, yielding a French-to-English phonetic continuum. Judgments switched from predominantly “French” to “English” responses when most of the French /R/s had been replaced by the English /J/. These findings suggested that listeners in the present study should have no difficulty identifying language identity in taco tokens spoken by Spanish and English monolinguals.

Another question of interest was whether the listeners' perceptual judgments would be based on all four segments in taco or on a single, prominent segment. For example, listeners might base their perceptual judgments on cross-language phonetic differences in /t/, for previous research has shown that listeners are very sensitive to the phonetic differences between prevocalic /p t k/ in Spanish and English (Flege, 1984; Flege & Hammond, 1982; Zuengler, 1985; see also Miller & Volaitis, 1989). Also, psycholinguistic research has shown that listeners are especially sensitive to the mispronunciation of word-initial consonants (Cole, 1981), that they may respond to syllable-initial consonants before listening to an entire syllable (Norris & Cutler, 1988), and that they can actively allocate attention to specific portions of a word (e.g., Pitt & Samuel, 1990a, 1990b; Terken, 1991). However, the production data obtained in Experiment 1 led us to hypothesize that the listeners would make use of acoustic information in all four segments of taco when judging language identity. If the bilingual talkers had not noted cross-language phonetic differences in all four segments of taco, it would be difficult to account for the interrelation in the extent to which they approximated English phonetic norms for the four segments of that word.
Method

The 42 taco tokens measured in Experiment 1 were presented to listeners over headphones at a comfortable level of about 74 dB SPL(A). Two listeners were native speakers of English. The third was a native speaker of German who had lived in the United States for several years. All three listeners were speech researchers who were familiar with Spanish-accented English and who had passed a pure-tone hearing screening at octave frequencies from 500 to 4000 Hz (re: 20 dB HL).

The listeners were first asked to decide if each word had been spoken in Spanish or in English. They signaled their choice by pushing a button marked “English” or “Spanish.” The words produced by the monolinguals were randomly presented 11 times each, with a 1.0-s interval between each response and the next word. Words spoken by the bilinguals were then presented in a similar fashion. The dependent variable was the number of “Spanish” responses given to the final 10 presentations of each word. Several days later the listeners rated the 42 taco tokens for goodness on a scale that ranged from a “good example of English taco” to “a good example of Spanish taco.” The words were randomly presented six times each in a single block. To signal their judgments, the listeners positioned the lever on a response box at some place along the scale, which yielded a number ranging from 1 to 256. The dependent variable was the mean of the last five ratings.

Results

Identification Data. Figure 1 shows the mean number of times the 42 taco tokens were identified as having been spoken in Spanish. Each mean was based on 30 forced-choice identifications (3 listeners x 10 presentations). The language identity of words spoken by the Spanish and English monolinguals (top panel) was identified correctly in nearly every instance. The language identity of taco tokens spoken in Spanish and English by bilinguals was not always evident to the listeners, however. The Spanish and English tokens produced by five early learners (#3, #5, #6, #7, #2) were differentiated by the listeners, whereas the remaining two early learners (#1, #4) seemed to produce taco in either a Spanish or an English mode. From the listeners' perspective, all seven late learners produced a small difference between Spanish and English taco tokens.

The number of “Spanish” judgments was examined in an ANOVA with six levels for the Group factor (one level each for the two monolingual groups, and two levels each for the two bilingual groups). The ANOVA was highly significant, $F(5, 36) = 33.5, p < .01$. A Newman-Keuls post-hoc test revealed that, as expected, the Spanish monolinguals' tokens were identified as having been produced in Spanish significantly more often than the taco tokens of the English monolinguals' (99 vs. 3%) ($p < .05$). Also, the tokens produced in Spanish by the two groups of bilinguals received significantly more Spanish identifications than the tokens they produced in English (early learners: 58 vs. 20%; late learners: 89 vs. 65%) ($p < .05$). A second ANOVA indicated that all three listeners performed the task in a comparable fashion. 6
The Word Unit in L2 Speech

Figure 1. Mean number of "Spanish" identifications of taco tokens produced (a) by Spanish and English monolinguals, (b) in both Spanish and English by early L2 learners, and (c) in both Spanish and English by late L2 learners.

Rating Data. As seen in Figure 2, much the same pattern of results was obtained in the rating experiment. This figure shows the mean ratings of 42 tokens of taco produced in Spanish or English by monolingual speakers of English (top), by early learners in Spanish and English (middle), and by late learners in Spanish and English (bottom). Each mean was based on 15 ratings (3 listeners x 5 replicate judgments). The effect of Group was significant, $F(5, 36) = 26.9, p < .01$. As expected, a post-hoc test revealed that the Spanish monolinguals' words were rated as significantly more Spanish-like than those of the English monolinguals (219 vs. 2) ($p < .05$). The early learners' renditions of taco in Spanish were rated as significantly more Spanish-like than their renditions of taco in English (167 vs. 81) ($p < .05$). A similar trend was evident for the late learners' renditions of taco in Spanish and English (217 vs. 185), but the difference was not significant ($p > .05$).

Multiple Regression. The identification and rating data differed in several ways. The late learners' Spanish and English taco tokens differed significantly in the
Figure 2. Mean ratings of tokens of *taco* produced (a) by Spanish and English monolinguals, (b) in both Spanish and English by early L2 learners, and (c) in both Spanish and English by late L2 learners.

identification but not the rating task. For three bilinguals (viz., early learners #5 and #7, late learner #1) Spanish and English *taco* differed in the identification but not the rating task. For two other bilinguals (early learner #4, late learner #4), the opposite held true. These differences suggested that the listeners may have responded to different acoustic phonetic information in the two tasks.

Two forward stepwise regression analyses were carried out to explore differences in the two perceptual tasks. One examined the relation between 17 acoustic measurements obtained in Experiment 1 and the percentage of "Spanish" responses in the identification task. The other examined the relation between the acoustic measurements and mean ratings. The regression analyses are presented in Table 4. The regression models accounted for a large amount of variance in both the identification and rating data (81 and 97%, respectively), which demonstrated that acoustic measurements made in Experiment 1 were appropriate. In the analysis of identification data, the VOT of /t/ accounted for 54% of the variance at Step 1 (later re-
The Word Unit in L2 Speech

### Table 4. Summary of forward stepwise regression analyses examining the relation between acoustic variables in *taco* tokens and listeners' perceptual judgments

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>Adjusted $R^2$</th>
<th>Change in $R^2$</th>
<th>$F$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Identification Scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>VOT of /t/</td>
<td>53.6</td>
<td>—</td>
<td>79.77</td>
<td>.000</td>
</tr>
<tr>
<td>2</td>
<td>Vowel 2 duration</td>
<td>64.4</td>
<td>10.8</td>
<td>16.91</td>
<td>.000</td>
</tr>
<tr>
<td>3</td>
<td>F2 Vowel 1</td>
<td>68.1</td>
<td>3.7</td>
<td>6.63</td>
<td>.017</td>
</tr>
<tr>
<td>4</td>
<td>F2 onset Vowel 2</td>
<td>70.4</td>
<td>2.3</td>
<td>4.43</td>
<td>.047</td>
</tr>
<tr>
<td>5</td>
<td>VOT of /t/ REMOVED</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>6</td>
<td>F1 Vowel 1</td>
<td>73.2</td>
<td>2.8</td>
<td>4.19</td>
<td>.053</td>
</tr>
<tr>
<td>7</td>
<td>F3 (onset) of Vowel 2</td>
<td>77.8</td>
<td>4.6</td>
<td>6.81</td>
<td>.016</td>
</tr>
<tr>
<td>8</td>
<td>F3 (offset) of Vowel 2</td>
<td>80.6</td>
<td>2.8</td>
<td>4.30</td>
<td>.049</td>
</tr>
<tr>
<td>Mean Ratings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>VOT of /t/</td>
<td>86.6</td>
<td>—</td>
<td>870.45</td>
<td>.000</td>
</tr>
<tr>
<td>2</td>
<td>VOT of /k/</td>
<td>89.6</td>
<td>3.0</td>
<td>33.09</td>
<td>.000</td>
</tr>
<tr>
<td>3</td>
<td>F3 of Vowel 1</td>
<td>90.9</td>
<td>1.3</td>
<td>14.63</td>
<td>.001</td>
</tr>
<tr>
<td>4</td>
<td>F2 (onset) of Vowel 2</td>
<td>92.6</td>
<td>1.7</td>
<td>17.68</td>
<td>.000</td>
</tr>
<tr>
<td>5</td>
<td>F1 (offset) of Vowel 2</td>
<td>95.2</td>
<td>2.6</td>
<td>24.00</td>
<td>.000</td>
</tr>
<tr>
<td>6</td>
<td>Vowel 2 intensity</td>
<td>96.8</td>
<td>1.6</td>
<td>14.81</td>
<td>.001</td>
</tr>
<tr>
<td>7</td>
<td>Vowel 2 duration</td>
<td>97.2</td>
<td>0.4</td>
<td>3.98</td>
<td>.059</td>
</tr>
</tbody>
</table>

Note: The $F$ values are tests of significant $R^2$ changes using the error term from the final step of the analyses.

moved), and the duration of Vowel 2 accounted for an additional 11% more of variance at Step 2. Five variables entered at Steps 3–7 accounted for a total of less than 5% of additional variance.\(^7\) In the analysis of the rating data, the VOT of /t/ accounted for 87% of the variance at Step 1, and six additional variables entered at subsequent steps accounted for less than 5% of additional variance. Taken together, these results suggest that the VOT of word-initial /t/ had a greater effect on ratings than language identification judgments but that, for both perceptual tasks, it was an important determinant of the listeners' perceptual judgments.

The conclusion concerning the effect of listening task on VOT perception was supported by two additional analyses. First, tests were carried out to compare the strengths of correlation between acoustic measurements and the two sets of perceptual data. The eight acoustic variables included in this analysis were those that had shown a significant correlation with both the identification and rating data (viz., VOT of /t/, Vowel 2 duration, the ratio of Vowel 1 and Vowel 2 duration, /k/ closure duration, Vowel 2 intensity, F2 of Vowel 1, and F2 frequency at both the onset and offset of Vowel 2). In only one instance did the strength of correlations between an acoustic variable and the two sets of perceptual data differ significantly. The VOT of /t/ showed a significantly stronger simple correlation with the mean ratings than with the percentage of Spanish identifications, $r = -.904$ versus .732; $M = 8.78, p < .01$.

Second, the nine acoustic variables identified as significant predictors of percep-
tual responses in earlier stepwise regression analyses were entered as forced variables into two new regression analyses, which accounted for 80% of the variance in the identification data, $F(9, 19) = 14.9$, and 97% of variance in the mean ratings, $F(9, 19) = 89.1; p < .01$. The variables having the three largest beta weights in the rating analysis were the VOT of /t/ ($-.565$), the onset frequency of F2 in Vowel 2 ($-.0.289$), and the offset frequency of F1 in Vowel 2 ($2.83$). Those having the largest beta weights in the identification analysis, on the other hand, were the onset frequency of F3 in Vowel 2 ($-.436$), the F2 frequency of Vowel 1 ($-.393$), and the F2 onset frequency of Vowel 2 ($3.78$). The beta weight of the VOT of /t/ ($0.075$) ranked eighth of nine variables in the analysis of identification scores and was nonsignificant ($p > .10$).

**Discussion**

Experiment 3 demonstrated that purely phonetic differences distinguishing the Spanish and English renditions of *taco* suffice to cue language identity. It also showed that the early learners produced larger phonetic differences in *taco* than did the late learners when switching from Spanish to English. One unexpected finding was that the bilinguals tended either to produce a fairly clear distinction between *taco* in Spanish and English or to produce *taco* in a Spanish-like or an English-like fashion in both languages. This is reminiscent of a finding reported by Lambert, Havelka, and Gardner (1959), who had French/English bilinguals read a list containing French and English words as rapidly as possible. Some words on their list, such as *chance* and *important*, exist in both French and English. Such words tended to be produced in either a French-like or English-like fashion.

It may be that bilinguals choose volitionally the phonetic characteristics of cognate words (see Romaine, 1989, for insightful discussion). If so, their production of a cognate word in a second language would not reflect their underlying phonetic competence in the second language unless they were actually attempting to produce the word like native speakers of the target second language. Several early learners in the present study managed to produce a Spanish-like *taco* in Spanish and an English-like *taco* in English. These individuals, we believe, intended to produce a cross-language difference. Their success in doing so demonstrated that they were cognizant of phonetic-level differences between the Spanish and English renditions of *taco*.

A regression analysis indicated that acoustic information in the first and last segments of *taco* (viz., VOT of /t/ and duration of Vowel 2) had an important influence on listeners' language identifications. However, a regression analysis of the listeners' ratings identified only the VOT of /t/ as an important predictor. The fact that vowel quality was not identified as an important predictor of the listeners' identifications or ratings came as a surprise, for vowel quality has been identified as an important predictor of intelligibility in studies of disordered speech (e.g., Ansel & Kent, 1992; Monsen, 1978). To determine whether the regression analyses may have underestimated the importance of vowel quality, we carried out new forward stepwise regression analyses that examined only the vowel formant frequency vari-
ables. These new analyses accounted for nearly as much variance in the identification and rating data (78 and 80%, respectively) as the ones reported earlier. This might mean that our listeners actually made use of acoustic cues distributed across the entire word in making perceptual judgments, an inference to be tested further in Experiment 4.

There are several possible reasons why the VOT of the word-initial /t/ in *taco* had an important effect on the listeners' judgments. Listeners are more sensitive to mispronunciation of word-initial than word-final or intervocalic consonants (Cole, 1981). They may be especially sensitive auditorily to the voicing feature in prevocalic stop consonants. Flege (1984) found that native English listeners could detect small differences in /t/s produced by native and French speakers of English in a paired-comparison task. Flege and Hammond (1982) found that native English speakers accurately reproduced nonnative characteristics of word-initial /t/s in a delayed mimicry task (see also Zuengler, 1985). In a goodness rating task, English listeners display a clearly specified range of preferred VOT values in prevocalic /p t k/ (e.g., Miller & Volaitis, 1989).

Finally, the results of Experiment 3 suggested that the listeners may have made greater use of the VOT of /t/ in the identification than rating task. It is important to recall, however, that all three listeners participated in the identification task prior to the rating task. The identification task may therefore have constituted training for the rating task (Whalen, Abramson, Lisker, & Mody, 1990). Also, the three listeners all had training in speech research, and all were highly familiar with Spanish-accented English. Their results might therefore not generalize to other, untrained listeners. These possibilities were tested in the final two experiments.

**EXPERIMENT 4**

Experiment 4 examined native English speakers' perceptual sensitivity to acoustic differences distinguishing the Spanish and English renditions of *taco*. Three acoustic dimensions were manipulated systematically in a synthetic continuum that ranged from a “Spanish” to an “English” *taco*: VOT of word-initial /t/, spectral quality of the first vowel, and duration of the second vowel. The use of synthetic speech stimuli permitted us to study the effect of each dimension independently of variations along the other two dimensions.

Previous research has suggested that rating and identification responses provide variant measures of the same perceptual processes (e.g., Leather, 1986; Repp, 1984). We tested this by assigning one-half of the 30 untrained listeners recruited for the present experiment to an identification task and the other half to a rating task. In the identification task, subjects identified synthetic *taco* tokens as having been produced in English or Spanish. In the rating task, subjects evaluated the same stimuli for degree of Englishness. We wanted to learn if the VOT manipulation would have a greater effect on the ratings than on the language identifications. We also wanted to determine whether all three dimensions would have a significant effect on the listeners' perceptual judgments and, if so, whether the relative importance of the three dimensions in the two perceptual tasks would differ.
Method

Stimuli. A formant synthesizer was used to generate a 63-member Spanish to English taco continuum. Selection of the three acoustic phonetic dimensions that were manipulated in synthesis was based on the Experiment 1 production results and on the results obtained in a study by Zuengler (1985) in which 45 native Spanish adults were asked to speak with a feigned "American" accent. Ninety-one percent of these talkers were judged to have modified their production of Spanish /t/ by making it more English-like. Many also modified their Spanish /o/ (by lengthening it and/or increasing degree of diphthongization) and many modified the spectral quality of Spanish /a/. The synthesis parameters associated with the word-initial /t/, Vowel 1 quality, and Vowel 2 duration in taco were manipulated in the present study so that they ranged from values appropriate for English to values appropriate for Spanish. The "Spanish" and "English" values were the average values obtained for four male Spanish and four male English monolinguals from Experiment 1. All other parameters were set to "neutral" values, defined as the mean observed for all eight monolinguals.

VOT of the word-initial /t/ ranged in approximately 7.5-ms steps from 20 ms (Step 1) to 65 ms (Step 7). The spectral quality of Vowel 1 ranged in three steps. The frequencies of the first three formants in the Spanish endpoint stimuli (Step 1) were 695, 1409, and 2367 Hz, respectively. The F1–F3 frequencies of the neutral (Step 2) stimuli were 639, 1308, and 2274 Hz. The frequencies of the English (Step 3) endpoint stimuli were 600, 1205, and 2225 Hz. Vowel 2 duration was also varied in three steps from 70 ms (the Spanish endpoint value) to 104 ms (the neutral value) to 146 ms (the English value) at Step 3. For the most part, the subjects who took part in a pilot experiment heard the expected four segments in the synthetic taco stimuli.

Subjects. The 30 subjects who participated in this experiment were recruited through advertisements on bulletin boards and in the university newspaper. Most were students or staff members at the University of Alabama at Birmingham. Fifteen subjects each were randomly assigned to one of two perceptual tasks. All subjects were monolingual native speakers of American English who had passed a pure-tone hearing screening at octave frequencies from 500 to 4000 Hz (re: 25 dB HL). A language background questionnaire indicated that the two groups were fairly homogenous.

Procedure. The synthetic stimuli were bandpass filtered (80–4800 Hz) before being presented binaurally via headphones. The subjects were tested one at a time in a sound booth. To help orient the subjects, we played a recording of the taco tokens that had been produced by the Spanish and English monolinguals from Experiment 1. These tokens were presented in separate blocks before the experiment began. The subjects were then told that they would hear computer-generated tokens of "Spanish" and "English" taco. The subjects assigned to the identification task were told to push a button marked "Spanish" or "English," depending on whether each
The Word Unit in L2 Speech

401

token was more Spanish-like or English-like. Those assigned to the rating task were
told to type a number ranging from 1 ("least English-like") to 9 ("most English-like")
on a keyboard. No feedback was provided to the subjects in either group.

The 63 stimuli were randomly presented eight times each to both groups of
subjects, with a 1.0-s delay between each response and presentation of the next
stimulus. The dependent variable for the identification task was the number of times
each of the 63 synthetic stimuli was identified as an "English" rendition of *taco.*
There were eight random presentations of each stimulus, so the identification task
scores ranged from 0 to 8. One was subtracted from mean ratings obtained for each
of the 63 stimuli, so the mean ratings also ranged from 0 to 8.

**Results**

*Identification and Rating Data.* The effect of manipulating the three acoustic
parameters on listeners' judgments are shown in Figure 3. The top panel shows the
effect of VOT, averaging across the variations in Vowel 1 quality and Vowel 2
duration. The middle panel shows the effect of Vowel 1 quality (averaging over the
other two parameters), and the bottom panel shows the effect of Vowel 2 duration
(again, averaging over the other two parameters). It can be seen that the manipulation
of all three acoustic phonetic dimensions affected the listeners' language identifi-
cations and goodness ratings. The frequency of English judgments decreased from
an average of 5.0 (62%) for the stimuli with the longest VOT value to 3.1 (39%) for
stimuli with the shortest VOT value. The average number of English judgments
decreased from an average of 5.1 (64%) for stimuli with the English-like Vowel 1
formant frequencies to 3.2 (40%) for stimuli with the Spanish-like formant frequen-
cies. The English judgments decreased from an average of 5.7 (71%) for stimuli with
Vowel 2 durations appropriate for English to 2.4 (30%) for the stimuli with the
Spanish-like Vowel 2 durations. Much the same pattern of results was evident for
the mean ratings, although Vowel 2 duration seemed to exert a larger effect on
identifications than ratings.

The scores obtained in the two perceptual tasks were submitted to an ANOVA in
which Task (identification vs. rating) served as a between-subjects factor and VOT
(7 levels), Vowel 1 quality (3 levels), and Vowel 2 duration (3 levels) served as
repeated measures. As expected, the manipulation of all three acoustic phonetic
dimensions exerted a significant effect ($p < .01$) on the listeners' judgments, VOT:
$F(6, 168) = 63.3$; Vowel 1 quality: $F(2, 56) = 63.4$; Vowel 2 duration: $F(2, 56) =
62.3$. Also as expected, the Task $\times$ Vowel 1 quality and Group $\times$ VOT interactions
were nonsignificant, $F(2, 56) = .40$, $F(6, 168) = .307, p > .10$, whereas the Task $\times$
Vowel 2 duration interaction did reach significance, $F(2, 26) = 6.16, p < .01$.

*Regression Analyses.* The identification and rating scores were submitted to
separate forward stepwise multiple regression analyses. These analyses examined
the effects of the three acoustic dimensions on the listeners' perceptual judgments.
The regression model accounted for a significant 49% (adjusted $R^2$) and 57% of
variance in the identification and rating data, respectively ($p < .01$). The results of these analyses agreed with the ANOVA in suggesting that Vowel 2 duration had a greater effect on responses obtained in the language identification than rating task. In the identification analysis, Vowel 2 duration accounted for 31% of variance at Step 1, Vowel 1 quality accounted for an additional 11% of variance at Step 2, and VOT accounted for an additional 7% of variance at Step 3. In the rating analysis, Vowel 1 quality accounted for 25% of the variance at Step 1, VOT accounted for 16% more variance at Step 2, and Vowel 2 duration accounted for 16% more variance at Step 3.
Discussion

The results of this experiment indicated that untrained listeners assigned to a language identification task, as well as those assigned to a goodness rating task, made use of information distributed over the entire word. Listeners showed systematic effects of the VOT of /t/, Vowel 1 quality, and Vowel 2 duration in judging the members of a synthetic Spanish-to-English taco continuum. This was by no means the only possible outcome, for the subjects assigned to either task might have narrowly focused attention on a single acoustic phonetic dimension (Norris & Cutler, 1988; Pitt & Samuel, 1990a, 1990b; Terken, 1991). The perceptual findings obtained here therefore parallel the production results obtained in Experiment 1. There it was shown that the extent to which Spanish/English bilinguals approximated English phonetic norms for the four segments in taco were interrelated.

Subjects assigned to the identification and rating conditions of the present experiment showed similar effects for two of the acoustic phonetic dimensions examined (VOT of /t/, Vowel 1 quality) but differed for a third (Vowel 2 duration). Perhaps subjects in the identification task allocated more attention to Vowel 2 duration than did subjects in the rating task. Alternatively, the identification subjects may have been influenced to a greater extent by rhythmic properties of the stimuli than were the rating subjects. The first syllable of the taco tokens produced by both the English and Spanish monolinguals in Experiment 1 received greater stress than the second syllable. Despite this, the second vowel was longer than the first vowel in words spoken by the English monolinguals (149 vs. 118 ms) but not the Spanish monolinguals (63 vs. 113 ms) because utterance-final vowels are lengthened to a greater extent in English than Spanish (Delattre, 1966; Oller, 1979; Pinkerton-Hutchinson, 1973). As a result, the Spanish and English renditions of taco had distinctly different rhythmic characteristics. This could not be apprehended by the listeners, of course, until they had heard the second vowel.

The results did not support the hypothesis (see Experiment 3) that listeners make greater use of variations in VOT of word-initial /t/ when rating words for goodness than when attempting to determine language identity. VOT was entered at Step 2 in the regression analysis of the mean ratings, accounting for just 16% of variance. It was entered at Step 3 in the analysis of identification data, accounting for only 7% of variance. Experiments 3 and 4 provided seemingly divergent results concerning listeners' overall perceptual use of VOT in the word-initial /t/ of taco. In Experiment 3, VOT was found to be the most important determinant of language identification and goodness judgments. In Experiment 4, it was less important than Vowel 1 quality and/or Vowel 2 duration.

One possible explanation for the apparently discrepant results of the two experiments is the use of natural versus synthetic speech stimuli. Perhaps VOT had a smaller overall effect on listeners' response here than in Experiment 3 because the synthesis did not represent the full range of temporal and spectral cues that accompany variations in VOT (Lisker & Abramson, 1971; Williams, 1979). Another possible explanation is that the Experiment 3 subjects were far more familiar with Span-
Figure 4. The mean number of identifications as "English" by the three subjects, each who were most or least familiar with Spanish-accented English. The brackets enclose ± 1.0 standard error.

ish-accented English than those in Experiment 4. A post-hoc analysis of the Experiment 4 data was carried out to assess the second possibility. We computed a "familiarity with Spanish" index for the 30 subjects who participated in Experiment 4. The index was based on the subjects’ estimates of (a) how often they had been exposed to Spanish or Spanish-accented English; (b) how many Spanish-speaking persons, if any, they knew; and (c) how many weeks, if any, they had spent in a Spanish-speaking country. The average difference in responses to Step 1 and Step 7 of the VOT manipulation was computed for the 15 rating and 15 identification subjects, then standardized. There was not a significant correlation between the standardized VOT effect scores and the familiarity-with-Spanish scores, $r(28) = .023, p > .10$.

Figure 4 shows the mean number of English identifications given in response to stimuli differing in VOT by the three subjects who were the "least familiar" with Spanish accents and the three subjects who were "most familiar." The least familiar subjects showed a somewhat larger effect of VOT than the most familiar subjects. The least familiar subjects therefore resembled the trained listeners of Experiment 3 to a greater extent than the most familiar subjects, which is just the opposite of what one would expect if familiarity accounted for the differing effects of VOT observed in Experiments 3 and 4.

EXPERIMENT 5

The apparently different effect of VOT in Experiments 3 and 4 might have resulted from differences in familiarity with Spanish-accented English by the listeners in those experiments, from the use of different stimuli (natural vs. synthetic realizations of taco), or from the order in which subjects participated in perceptual tasks. A post-hoc
The Word Unit in L2 Speech

Results and Discussion

The findings obtained here replicated the results of Experiment 3, where trained listeners judged native Spanish early learners of English to have produced more substantial English versus Spanish differences than native Spanish late learners of English. Listeners who were assigned to the language identification task gave more English responses to the early learners’ productions of English *taco* than Spanish *taco* (79 vs 39%). They also gave more English responses to the late learners’ productions of *taco* in English than Spanish (45 vs. 17%). The percentage of English responses were examined in a Group x Language ANOVA, which yielded a significant two-way interaction, $F(1, 14) = 14.5, p < .01$. This is because a larger English versus Spanish difference was evident for the early than late learners. Listeners assigned to the rating task gave higher (more English-like) ratings to the early learners’ English than Spanish *taco* tokens (6.7 vs. 3.6). They also gave higher ratings to the late learners’ English than Spanish *taco* tokens (4.1 vs. 2.5). A significant Group X Language interaction was obtained, $F(1, 14) = 44.5, p < .01$.

Correlation analyses examining the two sets of perceptual data and the acoustic measurements obtained in Experiment 1 are presented in Table 5. The two sets of correlations were generally similar. Of the 18 acoustic dimensions, just three were correlated significantly with mean identification scores (viz., VOT of word-initial /t/, closure duration of inter vocalic /k/, and Vowel 2 intensity), and four acoustic dimensions were correlated significantly with mean ratings (viz., VOT of /t/, duration of /k/, F2 of Vowel 1, and onset frequency of F2 in Vowel 2) ($p < .01$). Vowel 2 duration showed a slightly stronger correlation with the identification than rating data, $r = .421$ versus $r = .340$, but both correlations were nonsignificant ($p > .10$). This finding suggests that the greater use of Vowel 2 duration by subjects assigned to the identification task than by subjects assigned to the rating task of Experiment 4 may have been due to the use of synthetic stimuli. The correlation between the VOT of /t/ and the mean identifications and ratings obtained in the present experiment were quite similar, $r = .888$ and $r = .892$, respectively. This suggests that the apparently greater use of /t/ VOT in the rating than identification task of Experiment 3 may have been due to a fixed sequence of testing.

The five acoustic variables that accounted for the most variance in previous regression analyses (viz., VOT of /t/, VOT of /k/, Vowel 2 duration, F2 of Vowel 1, and onset frequency of F3 in vowel 2) were entered as forced variables in multiple
Table 5. Simple correlations between acoustic measurements of 28 Spanish and English *taco* tokens and the perceptual judgments obtained in Experiment 4

<table>
<thead>
<tr>
<th>Acoustic Variable</th>
<th>Identifications</th>
<th>Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>VOT of /t/</td>
<td>0.888*</td>
<td>0.892*</td>
</tr>
<tr>
<td>F1–Vowel 1</td>
<td>0.022</td>
<td>-0.074</td>
</tr>
<tr>
<td>F2–Vowel 1</td>
<td>-0.461</td>
<td>-0.548*</td>
</tr>
<tr>
<td>F3–Vowel 1</td>
<td>-0.320</td>
<td>-0.254</td>
</tr>
<tr>
<td>Vowel 1 duration</td>
<td>0.118</td>
<td>0.128</td>
</tr>
<tr>
<td>/k/ VOT</td>
<td>-0.287</td>
<td>-0.282</td>
</tr>
<tr>
<td>/k/ closure duration</td>
<td>-0.596*</td>
<td>-0.620*</td>
</tr>
<tr>
<td>/k/ burst intensity</td>
<td>-0.002</td>
<td>-0.173</td>
</tr>
<tr>
<td>F1–Vowel 2 (onset)</td>
<td>-0.175</td>
<td>-0.173</td>
</tr>
<tr>
<td>F1–Vowel 2 (offset)</td>
<td>-0.349</td>
<td>-0.281</td>
</tr>
<tr>
<td>F1–Vowel 2 movement</td>
<td>0.259</td>
<td>0.187</td>
</tr>
<tr>
<td>F2–Vowel 2 (onset)</td>
<td>0.429</td>
<td>0.539*</td>
</tr>
<tr>
<td>F2–Vowel 2 (offset)</td>
<td>0.059</td>
<td>0.157</td>
</tr>
<tr>
<td>F2–Vowel 2 movement</td>
<td>0.424</td>
<td>0.474</td>
</tr>
<tr>
<td>F3–Vowel 2 (onset)</td>
<td>0.032</td>
<td>-0.046</td>
</tr>
<tr>
<td>F3–Vowel 2 (offset)</td>
<td>0.165</td>
<td>0.015</td>
</tr>
<tr>
<td>F3–Vowel 2 movement</td>
<td>-0.137</td>
<td>-0.043</td>
</tr>
<tr>
<td>/o/ duration</td>
<td>0.421</td>
<td>0.343</td>
</tr>
<tr>
<td>/o/ intensity</td>
<td>-0.503*</td>
<td>-0.469</td>
</tr>
</tbody>
</table>

*Significance at the .01 level (two-tailed).

regression analyses. Analysis of the identification data accounted for 91% of variance, $F(5, 21) = 52.1, p < .01$. Analysis of the rating data accounted for 86% of variance, $F(5, 21) = 31.6, p < .01$. The VOT of /t/ and the F2 frequency of Vowel 1 were significant predictors of both the rating and identification scores. The VOT of /k/ was significant for the rating but not the identification data ($p < .01$). Perhaps listeners in the language identification task were able to reach a decision after hearing just the first syllable and did not attend to properties of the second syllable (including /k/). F3 onset frequency in Vowel 2 was nonsignificant in both analyses, as was Vowel 2 duration. This last finding came as a surprise, given the significant effect of Vowel 2 duration in the experiment with synthetic stimuli.

It is noteworthy that the VOT of /t/ had the highest beta weights in analyses of both the identification and rating data (.741 and .777, respectively). In forward, stepwise regression analyses, the VOT of /t/ was the only variable identified as a significant predictor of identification scores (accounting for 78% of variance), and it was the most important predictor of mean ratings (accounting for 79% of variance). These results agree with Experiment 3 in suggesting that the VOT of /t/ may be more salient than any other single acoustic dimension distinguishing Spanish from English *taco*. 
Production of the word *taco* was examined in the first two experiments of this study. Productions by Spanish and English monolinguals were found to differ systematically at a phonetic level. Although Spanish/English bilinguals differentiated *taco* when switching between Spanish and English, their differentiations were often incomplete; that is, they merely approximated English phonetic norms for the four segments of *taco* in many instances. An interrelationship was observed in the bilinguals' production of the four segments of *taco*. The more accurately one segment was produced as the bilinguals switched from Spanish to English, the more accurately the other segments were also produced. This finding argues against the common view that L1 sounds are "substituted" for sounds in L2 words. It is consistent, on the other hand, with the general approach of nonlinear phonology. According to this approach, the implementation of a phonetic segment depends on its place within a prosodic hierarchy, as defined by a metrical tree (Beckman & Edwards, 1992; Goldsmith, 1990).

The acoustic analyses suggested that early learners produced larger acoustic differences than did the late learners when switching from Spanish to English. The perception experiments verified this conclusion, for listeners heard larger Spanish versus English differences in words spoken by early than late learners. Although this came as no surprise (see, e.g., Flege, 1991; Flege & Fletcher, 1992), it was nevertheless surprising that some early learners' productions of English *taco* were often not fully English-like.

A follow-up production experiment suggested that Spanish/English bilinguals may produce /t/ with shorter—and thus more Spanish-like—VOT values in English words that have a counterpart in Spanish than in English words that are not related to a word in the Spanish lexicon. The results do not make clear whether this is a fixed limitation, or if it resulted from a volitional choice on the part of some talkers (i.e., an intent to say *taco* in an English sentence with Spanish-like phonetic characteristics). It was clear, however, that some of the early learners were fully capable of producing Spanish-like and English-like renditions of *taco*. This demonstrated that they were aware of many if not all of the phonetic differences distinguishing the four segments of Spanish *taco* from the four corresponding segments of their English rendition of *taco*.

Previous research has suggested that rating and identification tasks provide variant measures of the same perceptual processes (e.g., Leather, 1986; Repp, 1984). In the present study, we found that the frequency with which listeners identified various words as having been spoken in English, and listeners' ratings of the degree of "Englishness" of various tokens of *taco*, often corresponded closely. There were two exceptions to this general rule, however. In Experiment 3, word-initial VOT had a greater effect on ratings than language identifications, and in Experiment 4 the duration of the final vowel in *taco* had a greater effect on identifications than ratings. However, neither difference was replicated in Experiment 5, where untrained listeners were randomly assigned to evaluate naturally produced *taco* tokens in either a language identification or rating task.
Listeners who participated in the perceptual experiments of the present study were not called upon to make categorical judgments of segment identity (e.g., /t/ vs. /d/). Their judgments were based on subtle yet systematic cross-language phonetic differences. The listeners had no difficulty distinguishing tokens of taco that had been spoken by Spanish and English monolinguals. This demonstrated that purely phonetic differences distinguishing the Spanish and English renditions of taco can be detected and that they may suffice to cue language identity. We found that when hearing words spoken by bilinguals, listeners drew upon acoustic information associated with all four segments of taco. This finding need not have been obtained, for listeners are able to narrowly focus attention on specific phonetic segments within a word (Cole, 1981; Norris & Cutler, 1988; Pitt & Samuel, 1990a, 1990b; Terken, 1991).

Our native English listeners' tendency to make use of acoustic information distributed over the entire word paralleled the Spanish/English bilinguals' tendency to produce all segments in a word relatively accurately or inaccurately. Taken together, the results suggest that the word may represent an important unit in L2 speech acquisition. This view is consistent with the observation by Trudgill (1986, p. 58) that, during the acquisition of a second dialect, an attempt is made to "accommodate" to the speech of the target language community by modifying the pronunciation of particular words. It is also consistent with the observation that, when attempting to gauge the distance of nonsense words from the patterns acceptable for real English words, subjects make reference to the whole word, not just parts of the word (Ohala & Ohala, 1986). Finally, the results are consistent with the notion that, in L2 learning, changes in segmental articulation may diffuse across the lexicon of the L2 (Gatbonton, 1983).

Despite the convergent evidence just mentioned, the conclusions reached here concerning L2 production should nevertheless be regarded as tentative. The present study examined only one word, albeit in great detail. It will therefore be necessary to replicate the production findings reported here to additional words, including words that have cognates in the L1 as well as those that cannot be related to any word in the L1 lexicon.

(Received 22 February 1993)

NOTES

1. The one acoustic variable in Tables 1 and 2 that was not included in the correlation analyses was the duration ratio of Vowel 1/Vowel 2.

2. The vowel height effect arises from differences in the amount of time needed to reestablish the transglottal pressure drop required for vocal fold vibration to occur in word-initial stops that are coproduced with vowels differing in oral cavity volume.

3. In a study by Elman, Diehl, and Buchwald (1977), the categorization of word-initial stops with short-lag VOT by certain bilinguals varied as a function of language "set." Bilabial stops that had been produced with short-lag VOT values typical for Spanish /p/ and English /b/ were heard as /b/ in the English set, but as /p/ in the Spanish set.

4. This assumes, of course, that Spanish /a/ and English /a/ can be regarded as functionally the "same" phoneme.

5. Of course, this finding is no guarantee that listeners in the present experiment would be able to identify the language in which taco tokens had been spoken. The difference in the "r" sounds of French and
English might be regarded as a qualitative difference in sound type rather than a quantitative phonetic difference like the ones distinguishing Spanish and English inco.

6. The identification scores obtained from the bilingual subjects were submitted to a Group (early vs. late) x Listener (3 levels) x Language (Spanish vs. English) ANOVA. A significantly larger proportion of Spanish identifications was obtained for words spoken in Spanish than English (57 vs. 27%), F(1, 36) = 43.3, p < .01. The Group x Language interaction was nonsignificant, F(1, 36) = 2.07, p > .05, suggesting that both groups of bilinguals were able to produce a perceptible difference in taco when speaking Spanish and English. The Listener factor was nonsignificant, F(2, 36) = 1.64, p = .206, and did not interact significantly with the other two factors. This suggests that the native German listener performed the identification task much like the two native English listeners.

7. The VOT of /t/ was apparently removed at Step 5 because it was correlated with Vowel 2 duration, r = .500, /k/ duration, r = -.367, Vowel 2 intensity, r = .480, the F2 of Vowel 1, r = -.381, the onset F2 of Vowel 2, r = .741, and the offset F2 of Vowel 2, r = .564.

8. Three native English and two native Spanish speakers served as subjects in a pilot experiment in which the first and the second syllables were presented separately for open set identification. The pilot subjects always identified the initial stop as /t/, even those synthesized with the short-lag Spanish VOT value. This means that a systematic effect of VOT in Experiment 4 can probably not be attributed to a difference in phonemic perception (e.g., a change from /t/ to /d/). The pilot subjects usually heard the vowel of the first syllable as /a/ or /A/. They tended to hear the vowel of the second syllable as /o/. The medial /k/ was heard as /p/, /t/, or /k/ from a few instances, as having been omitted. However, when the two syllables were presented together as they had been synthesized, with transitions leading into and out of the medial consonant, an intervocalic /k/ was plainly evident.

9. As revealed by f tests, the identification and rating subjects did not differ significantly according to age (29 vs. 27 years), familiarity with Spanish (2.9 vs. 2.7 on a 7-point scale), time spent in a Spanish-speaking country on vacation (0.3 vs. 1.5 weeks), self-reported ability to imitate foreign accents or dialects (2.8 vs. 2.9), self-reported ability to recognize the language background of nonnative talkers (3.9 vs. 4.1), number of Spanish-speaking acquaintances (3.0 vs. 4.5), or self-estimated familiarity with Spanish-accented English (4.0 vs. 3.7).

10. A post-hoc test revealed that the responses obtained for all three levels of Vowel 1 quality (English, neutral, Spanish), and all three levels of Vowel 2 duration, differed significantly (p < .01). The VOT values at Step 4 (the neutral value) to Step 7 (the English value of 65 ms) differed significantly from the Spanish-like VOT values at Step 1 to Step 3, and Step 7 differed significantly from Steps 4 and 6 (p < .01).

11. Perhaps rhythmic properties of word stimuli are used to a greater extent in a language identification task, but only when stimulus variation is restricted, as was the case for our synthetic taco stimuli.

12. Only five variables were entered because responses to just 28 stimuli were being examined.

REFERENCES


