Production of the word-final English /t/-/d/ contrast by native speakers of English, Mandarin, and Spanish

James Emil Flege, Murray J. Munro, and Laurie Skelton
Department of Biocommunication, University of Alabama at Birmingham, Birmingham, Alabama 35294-0019

(Received 18 April 1991; accepted for publication 1 April 1992)

The primary aim of this study was to determine if adults whose native language permits neither voiced nor voiceless stops to occur in word-final position can master the English word-final /t/-/d/ contrast. Native English-speaking listeners identified the voicing feature in word-final stops produced by talkers in five groups: native speakers of English, experienced and inexperienced native Spanish speakers of English, and experienced and inexperienced native Mandarin speakers of English. Contrary to hypothesis, the experienced second language (L2) learners' stops were not identified significantly better than stops produced by the inexperienced L2 learners; and their stops were correctly identified significantly less often than stops produced by the native English speakers. Acoustic analyses revealed that the native English speakers made vowels significantly longer before /d/ than /t/, produced /t/-final words with a higher F1 offset frequency than /d/-final words, produced more closure voicing in /d/ than /t/, and sustained closure longer for /t/ than /d/. The L2 learners produced the same kinds of acoustic differences between /t/ and /d/, but theirs were usually of significantly smaller magnitude. Taken together, the results suggest that only a few of the 40 L2 learners examined in the present study had mastered the English word-final /t/-/d/ contrast. Several possible explanations for this negative finding are presented. Multiple regression analyses revealed that the native English listeners made perceptual use of the small, albeit significant, vowel duration differences produced in minimal pairs by the nonnative speakers. A significantly stronger correlation existed between vowel duration differences and the listeners' identifications of final stops in minimal pairs when the perceptual judgments were obtained in an "edited" condition (where post-vocalic cues were removed) than in a "full cue" condition. This suggested that listeners may modify their identification of stops based on the availability of acoustic cues.

PACS numbers: 43.70.Kv, 43.70.Fq, 43.71.Hw

INTRODUCTION

Adults who learn a second language (L2) often have difficulty producing consonants that are found in the L2 but not in their native language (L1). For example, Flege and Davidian (1984) transcribed the final stops in minimally paired English words ending in /b d g/ and /p t k/ that had been spoken by native speakers of English, Spanish, and Taiwanese. Most of the native speakers' stops were heard as intended. However, a few of the nonnative speakers' stops were perceived to have been omitted; and over one-third of their /b,d,g/ tokens were perceived to have been devoiced. The segmental production errors that were noted appear to have arisen from phonological differences between English and the nonnative subjects' L1. Taiwanese permits /p t k/ but not /b d g/ in word-final position (Cheng, 1968). Spanish is said to have a phonological rule which devoices phonologically voiced stops in utterance-final position. The few word-final stops that do occur in Spanish words tend to be spirantized or omitted (Harris, 1969).

According to the speech learning model (SLM) developed by Flege, adult learners of an L2 who have received sufficient native-speaker input can master "new" sounds in the L2 (Flege, 1988a, 1991, 1992a,b). By definition, new sounds are those L2 vowels and consonants that learners of an L2 regard as falling outside the L1 phonetic inventory. According to the SLM, sounds classified as "similar" cannot be mastered because they continue to be identified perceptually in terms of a phonetically distinct counterpart in the L1 (e.g., the aspirated /t/ of English vis-à-vis the unaspirated /t/ of Spanish). Equivalence classification is said to prevent L2 learners from establishing phonetic categories for similar L2 vowels and consonants, which limits the accuracy with which these L2 sounds can be produced (Flege, 1987). The primary purpose of this study was to test the SLM's hypothesis regarding the production of new L2 consonants. More specifically, we wanted to learn if it is possible for adults whose L1 lacks a contrast between word-final /t/ and /d/ to accurately produce /t/ and /d/ in the final position of English words.

The two groups of native Spanish subjects examined in the present study differed according to length of residence in the US. All of these subjects were living in a predominantly English-speaking environment at the time of the study; most were affiliated with a large urban medical center. If subjects in the more experienced group had received sufficient L2
input, the SLM would predict their mastery of English /t/ and /d/. However, given continuing uncertainty as to what constitutes "sufficient" L2 input, a more conservative prediction that one might derive from the SLM is that only some of the experienced Spanish subjects would master English /t/ and /d/, although all of the experienced subjects should resemble the naive English speakers to a greater extent than subjects in the inexperienced Spanish group.

Both predictions were based on an important assumption about how the Spanish subjects would perceive word-final English stops, viz., that the experienced Spanish speakers of English would not identify word-final English /t/ and /d/ tokens in terms of a pre-existing Spanish stop category. Such an assumption might be questioned, however. The phonemes /p/, /k/, /h/, /v/, and /g/ do not occur in the final position of native Spanish words, but they do occur in some loanwords and proper names of Catalonian origin. Moreover, some native Spanish words such as novedad ("novelty") or ciudad ("city") end in /d/, although this phoneme is normally realized as a voiced or voiceless fricative. It is conceivable, therefore, that experienced Spanish speakers of English might persist in identifying word-final realizations of English /d/ (and possibly /t/) in terms of a Spanish stop category. If so, the SLM would predict inaccurate, Spanish-like realizations of English /d/ and /t/ by even the experienced Spanish subjects.

We did not assess the interlingual identification of L1 and L2 stops in the present study, which means that differences between the native English and experienced Spanish subjects, should any be observed, would be difficult to interpret. As a safeguard, therefore, we examined two groups of native Mandarin subjects. These subjects would not be expected to identify word-final tokens of either English /t/ or /d/ with word-final stops in their L1 because Mandarin does not permit any obstruent consonants—voiced or voiceless—to occur in word-final position (Cheng, 1973; Howie, 1976; Norman, 1988). If new L2 consonants can indeed be learned by adults as hypothesized by the SLM, one would expect the experienced Mandarin subjects to outperform relatively inexperienced Mandarin subjects, and also to closely match native speakers of English in producing word-final /t/ and /d/.

Based on previous research, we expected the native speakers of English to make vowels substantially longer before /d/ than /t/, to sustain closure longer for /t/ than /d/, to sustain glottal pulsing longer in the closure interval of /d/ than /t/, and to produce words ending in /t/ with a higher first formant (F1) offset frequency than words ending in /d/ (Peterson and Lehiste, 1960; House, 1961; Suen and Beddoes, 1974; Wolf, 1978; Repp, 1978; Port, 1979, 1981; Hogan and Roszypal, 1980; Raphael, 1981; Hillenbrand et al., 1984; Lisken, 1986). The predictions stated above were tested by comparing acoustic measurements of stops produced by the native speakers of English and the two experienced L2 groups, and by comparing stops produced by the relatively experienced and inexperienced L2 learners.

Predictions of the SLM were also tested by having native English-speaking listeners identify the voicing feature in final stops produced by the native and nonnative speakers. Previous research suggested that all four acoustic dimensions mentioned above would influence the native English listeners' judgments of stops produced by fellow native speakers of English. This was not tested, however, because there was little variability in the data obtained from the native English speakers, whose /t/'s and /d/'s were identified correctly in most instances. An aim of the study was to assess the perceptual importance of the four acoustic phonetic distinctions produced by the nonnative speakers, whose stops were frequently misidentified. Vowel duration, which serves as a perceptual cue to the voicing feature in word-final stops produced by native speakers of English in utterance-final or prepausal position, was of special interest (e.g., Raphael, 1972; Repp, 1978; Wardrip-Fruin, 1982; Hogan and Roszypal, 1980; Port and Dalby, 1982). Crowther and Mann (1990) recently measured vowel duration and F1 offset frequency in English words spoken by native speakers of Mandarin and Japanese. These subjects produced a substantially smaller vowel duration distinction in an English minimal pair (pot versus pod) than did native speakers of English. (The nonnative speakers also produced a smaller F1 offset frequency difference than did the native English speakers.) The relatively small "voicing effect" produced by the nonnative speakers was expected from previous studies showing that the effect of stop voicing on preceding vowel duration is larger in English than in many other languages (e.g., Chen, 1970; Flege and Port, 1981; Stevens et al., 1986) and from studies showing that adult learners of English tend to produce smaller voicing effects than native speakers of English (Sioumi, 1976; Elsendoorn, 1980, 1982; Flege and Port, 1981; Mittleb, 1981a,b; Mack, 1982; Port and Mittleb, 1983; Lowie, 1988; Hiramatsu, 1990). The small voicing effects produced by Crowther and Mann's nonnative subjects might account, at least in part, for why roughly half of their final /d/ tokens were heard as /t/. This is by no means certain, however, because the misidentifications of /d/ may have been due to acoustic dimensions other than vowel duration, such as F1 offset frequency, closure voicing, or closure duration.

The perceptual efficacy of acoustic cues to the /t/-/d/ contrasts produced by the nonnative speakers was assessed in multiple correlation analyses. These analyses examined acoustic measurements of minimally paired CVC words produced by the nonnative speakers and native English listeners' identifications of final stops in those CVCs. It was hoped that these analyses would provide insight into how listeners process phonetic contrasts that do not fully conform to the phonetic norms of their native language. It is known that foreign-accented speech becomes more intelligible as listeners gain familiarity with it (Wingstedt and Schulman, 1984), perhaps because the listeners develop "correspondence rules" relating incorrectly implemented sounds onto the intended target-language phonetic categories. Here, we tested the hypothesis that listeners are capable of modifying their use of acoustic cues in phonetic segments produced by nonnative speakers. CVCs were presented in two listening conditions. In the "full cue" condition, the native English-speaking listeners heard unmodified versions of the CVC words. In an "edited" stimulus condition, any energy pres-
ent in the final stop closure interval was removed along with the final release burst (if one was present). We expected lower rates of correct identification for stops in the edited condition than in the full-cue condition (Malecòt, 1958; Ohde and Sharf, 1977; Piceny et al., 1986; Feige, 1989; Fege and Wang, 1990). If listeners are able to modify segmental perception, the correlation between the size of vowel duration differences in minimal pairs produced by the non-native speakers and the native English listeners' identification of final stops in those minimal pairs should be significantly stronger in the edited than full-cue condition because fewer acoustic cues to the /t/-/d/ contrast were available in the edited condition.

I. METHOD

Native and nonnative talkers produced minimally paired CVC words. The CVCs were low-pass filtered at 4.8 kHz and digitized at 10.0 kHz with 12-bit amplitude resolution. After preparation (see below), half of the CVCs were presented to native English-speaking listeners who identified word-final stops as /t/ or /d/. All of the CVCs were examined acoustically; then the acoustic measures were related in correlational analyses to the listeners' judgments of CVCs for which both acoustic and perceptual data were available.

A. Talkers

Five males and five females in each of five groups, all of whom were recruited through advertisements in the university newspaper or personal contacts, participated as talkers. All of these subjects were living in Birmingham, AL, at the time of the study. Subjects in the native English group were monolingual native speakers of American English with a mean age of 28 years. Five were from Alabama, and one each was from South Dakota, Missouri, Washington, DC, Ohio, and Tennessee. The native Spanish subjects came from several countries in Central and South America (Colombia—9, Nicaragua—2, Puerto Rico—2, Chile—2, Bolivia—1, Honduras—1, Venezuela—1, Guatemala—1, Argentina—1). Those in the “experienced” group had lived in the US longer than those in a relatively “inexperienced” group (9.0 vs 0.4 years on the average). Half of the native Mandarin speakers were from mainland China, the rest from Taiwan. The “experienced” Mandarin subjects had lived in the US longer than those in the “inexperienced” group (5.5 vs 0.9 years). A native Mandarin research assistant verified that all of the Chinese subjects, including those from Taiwan, spoke Mandarin as their native language.

All of the nonnative subjects had learned English as an L2 in adulthood. The relatively experienced and inexperienced L2 learners differed little in chronological age and amount of formal education in English. However, as summarized in Table I they differed somewhat according to other factors that could potentially have influenced their ability to pronounce English. Compared to the inexperienced Spanish subjects, the experienced native Spanish subjects had arrived in the US at a slightly earlier age (20 vs 26 years) and estimated speaking English slightly more often on a daily basis (75% vs 68%). Compared to the relatively inexperienced Mandarin subjects, the experienced Mandarin subjects had arrived in the US at a slightly earlier age (23 vs 27 years) and estimated speaking English slightly more often (51% vs 32%). We assumed, however, that differences between the experienced and inexperienced L2 learners, should any be observed, could be attributed primarily to differences in overall amount of English-language experience.

B. Speech stimuli

Each talker read four lists in which 34 randomized CVC words occurred at the end of the carrier phrase “I will say__.” All CVCs on each list contained either /t/-/d/ or /s/-/z/. (We blocked on vowel to minimizing spelling-induced mispronunciations.) Each list contained minimal pairs of the form /bVt/-/bVd/ and /sVt/-/sVd/ (e.g., beat-bead, sat-sad), which will be referred to as the /bVC/ and /sVC/ words, respectively. The members of minimal pairs were never immediately juxtaposed on the lists; and no words selected for analyses came from the very beginning or end of the lists. After being digitized, the 16 words spoken by each talker were edited from the carrier phrase and normalized for peak intensity.

C. Acoustic analyses

Four acoustic dimensions were measured in the /bVC/ and /sVC/ words. Vowel duration was measured from the first positive peak in the periodic portion of each waveform to constriction of the word-final /t/ or /d/. Constriction was defined as occurring when amplitude decreased abruptly and the waveforms became more nearly sinusoidal in shape. The changes in waveform shape indicated that frequencies in the glottal source above the region of F1 had been filtered out by closure of the vocal tract (Hillenbrand et al., 1984). Closure voicing was measured from the point of constriction, as

### Table I. Characteristics of the talkers in five groups who participated in the study. Each group had five male and five female subjects. Standard deviations are in parentheses.

<table>
<thead>
<tr>
<th>Subject group</th>
<th>AGE</th>
<th>YRS</th>
<th>AOA</th>
<th>LOR</th>
<th>PER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native English</td>
<td>27.9</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Inexperienced Spanish</td>
<td>26.7</td>
<td>6.1</td>
<td>25.7</td>
<td>0.4</td>
<td>68.0</td>
</tr>
<tr>
<td>Experienced Spanish</td>
<td>28.2</td>
<td>7.1</td>
<td>19.9</td>
<td>9.0</td>
<td>75.2</td>
</tr>
<tr>
<td>Inexperienced Mandarin</td>
<td>28.4</td>
<td>7.4</td>
<td>27.2</td>
<td>0.9</td>
<td>32.0</td>
</tr>
<tr>
<td>Experienced Mandarin</td>
<td>28.1</td>
<td>6.0</td>
<td>22.6</td>
<td>5.5</td>
<td>51.0</td>
</tr>
</tbody>
</table>

*AGE—chronological age, in years.
1YRS—years of classroom instruction in English.
1AOA—age upon arrival in the United States, in years.
1LOR—length of residence in the United States, in years.
1PER—self-estimates of percentage daily use of English.
just defined, to the last positive peak in any low-amplitude, sinusoidal voicing in the stop closure interval. There was no visible acoustic evidence of release of lingual constriction in 29 (7.3%) of the words ending in /t/ and 85 (21.2%) of the words ending in /d/. Stop closure duration could not be measured in these stops. In all others, duration of the closure interval was measured from constriction to the beginning of the release burst.

F1 offset frequency was measured from formant tracks such as those for three minimally paired English words shown in Fig. 1. These words were spoken by one of the native English subjects. As expected from previous research, the offset frequency of F1 was much the same in beat-bead and bit-bid, but it was higher in bat than bad. The formant tracks in each word were obtained by performing a series of LPC analyses using 14 coefficients for male talkers and 12 coefficients for female talkers. A 12.8-ms Hamming window was moved in 5.0-ms steps through each CVC waveform. A peak-picking algorithm identified the first five peaks in the smoothed LPC spectra. Peaks were then assigned by algorithm to formants. This last step was fully automatic for most tokens, but user intervention was needed for some tokens. For example, the formant tracks had to be hand edited when a peak that should have been assigned to F2 was incorrectly assigned to F1 because no peak in the F1 region was identified by the peak-picking algorithm. In other instances, a few spurious data points located between the F1 and F2 tracks, or between the F2 and F3, had to be removed by hand before analysis.

Three mean F1 values were obtained for each CVC, each representing a 15-ms segment containing frequencies that were obtained, in most instances, from three successive LPC analyses. The last of the three segments, which occurred at the end of the formant transitions leading into /t/ and /d/, was segmented from the following stop closure interval on the basis of frequency changes in the F2 and F3 tracks and large increases in formant bandwidths, which are not shown in Fig. 1. (Segmentation was based on changes in the shape and amplitude of time domain waveforms in a few tokens in which constriction location was not clear.) The frequency offset analysis, unlike the other acoustic analysis described thus far, was confined to the native speakers of English and the two experienced L2 groups because it was so time consuming.

D. Listening test

Copies of the /bVC/ words were edited as described by Flege (1989). Any energy present in the stop closure interval was removed along with the final release burst, if there was one. The edited stimuli therefore contained no acoustic information pertaining to lingual constriction or release. As noted earlier, not all stops had release bursts. We reasoned that the effect of editing would be larger for the stops with release bursts than for those without release bursts because more phonetically relevant information was removed from the CVCs that contained release bursts.

To provide an estimate of how many stops contained audible (as opposed to simply visible) release bursts, the first author judged the final stop in the 100 available tokens of beat and bead. These words were presented randomly over headphones to avoid biases. As summarized in Table II, the native English speakers were usually judged to have produced final /t/‘s with a voiceless release burst and to have produced final /d/‘s with a voiced release burst. Of the /t/‘s produced by the nonnative speakers, two (5%) were perceived to have no burst and 38 (95%) to have a voiceless release burst. Of the nonnative speakers' /d/‘s, six (15%)
TABLE II. Perceptual characteristics of final release bursts in *bead* and *bead* tokens as spoken by subjects in five groups. A listener judged final release bursts to be inaudible (none), voiced (+ voi), or voiceless (- voi).

<table>
<thead>
<tr>
<th></th>
<th>/t/</th>
<th>/d/</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>None</td>
<td>-Voi</td>
</tr>
<tr>
<td><strong>Native English speakers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>female</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td><strong>Inexperienced Spanish subjects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>female</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td><strong>Experienced Spanish subjects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>female</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td><strong>Inexperienced Mandarin subjects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>female</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td><strong>Experienced Mandarin subjects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>female</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

were judged to have no release burst, 13 (33%) to have a voiced release burst, and 21 (52%) to have a voiceless release burst. This suggested, paradoxically, that some of the non-native speakers' final /d/'s might be identified at higher rates in the edited than in the full-cue condition.

The /bVC/ words were presented to three male and seven female monolingual speakers of American English with a mean age of 28 years. Four were from Alabama and one each was originally from Tennessee, Mississippi, North Carolina, South Carolina, Massachusetts, and Maryland. Each listener passed a pure-tone hearing screening at octave frequencies ranging from 500 to 4000 Hz at 20 dB HL. They heard the CVC words over headphones at a comfortable level (about 74 dB SPL-A) while seated in a sound booth. The listeners were told to identify the final stop in each CVC as /t/ or /d/, and to guess if uncertain. Each CVC was presented 1 s after a response was received for the preceding CVC. The edited condition followed the full-cue condition after a short break. Eight blocks of 50 stimuli were presented in counterbalanced order within each half of the experiment. The blocks varied according to the vowel in the CVCs (/i/, /e/, /æ/) and the talkers' gender. For example, one block consisted of the *beat* and *bead* tokens produced by the 25 female talkers (5 groups × 5 talkers).

We determined how many of the ten native English listeners correctly identified the final stop in each word. Percent correct scores are reported in Sec. II. To correct for guessing and to avoid potential problems arising from heterogeneity of variance, statistical analyses were performed on A' values, a bias-free measure of perceptual sensitivity (Grier, 1971). Four A' scores were obtained for each talker, one for each minimal pair. The A' scores were based on the number of "hits" (correct identifications of intended /t/ as voiceless) and "false alarms" (incorrect identifications of intended /d/ as voiceless). Each A' score was based on 20 forced-choice judgments (10 listeners × 2 members of each minimal pair).

II. RESULTS

A. Listening test

The overall rate of correct identifications ranged from 68% for the final stops in *bit-bid* to 71% correct for the stops at the end of *bet-bed* and *bat-bad*. Rates were higher for /t/ than /d/ in both the full-cue (82% vs 65%) and edited conditions (72% vs 61%). This means that the /d/ stops were perceived to have been voiced more often than the /t/ stops were heard as voiced stops. The average rate of correct identifications dropped by 7% when closure voicing and release burst cues were removed in the edited condition. All listeners participated in the edited condition after the full-cue (unedited) condition, which may have helped them to identify stops in the edited condition. If so, the fixed order of the two conditions may account for why the removal of closure voicing and release burst cues had a somewhat smaller effect in the present study than in some previous studies (e.g., Wolf, 1978; Revoile et al., 1982; but see Malricot, 1958). The effect of editing was somewhat greater for /t/ than /d/ (10% vs 4%). It was roughly the same for the native speakers of English (7%) as it was for the experienced and inexperienced Spanish subjects (8%, 11%) and for the experienced and inexperienced Mandarin subjects (5%, 6%). The native English speakers' stops were identified correctly more often in the full-cue condition (95% correct) than the experienced Spanish and Mandarin subjects' stops (71%, 65%). The native speakers' stops were also identified correctly more often than those of the experienced Spanish and Mandarin subjects (73%, 62%).

The A' scores were submitted to a mixed-design group X minimal pair X condition ANOVA with repeated measures on the last two factors. The mean A' scores shown in Fig. 2 have been averaged over the four minimal pairs because the minimal pair factor was nonsignificant [F(3,135) = 0.249, p>0.10]. As expected, the listeners were significantly more sensitive to the voicing feature of stops presented in the full-cue than edited condition (0.803 vs 0.706) [F(1,45) = 41.2, p < 0.001]. The group X condition interaction was nonsignificant [F(4,45) = 1.34, p>0.10], suggesting that the talkers in all five groups produced perceptually useful acoustic cues to the /t/-/d/ contrast in the closure interval and/or the final release burst.

The significant main effect of group [F(4,45) = 7.05, p<0.001] was followed up by post-hoc comparisons using Tukey's HSD test. These tests revealed that the native English speakers' stops had significantly higher A' scores (0.953) than the experienced and inexperienced Spanish subjects' stops (0.751, 0.720). The native English speakers also received higher A' scores than the experienced and inexperienced Mandarin subjects (0.668, 0.679). Neither the ex-
FIG. 2. Mean perceptual sensitivity (A') of native English-speaking listeners to /t/-/d/ contrasts in minimal pairs spoken by native speakers of English and four groups of adult L2 learners in two listening conditions. The data have been averaged over four minimal pairs. The brackets enclose ± one standard deviation for the ten subjects per group.

experienced and inexperienced Spanish subjects, nor the experienced and inexperienced Mandarin subjects, differed significantly (p > 0.10). The two Spanish groups had somewhat higher A' scores than the corresponding Mandarin groups, but these cross-language differences were nonsignificant (p > 0.10).

Figure 3 shows the average A' scores obtained for each of the 50 talkers whose final stop production was assessed. The mean scores represent an average of the scores obtained in four minimal pairs. The scores obtained for the native English subjects ranged from 0.906 to 1.000 (perfect sensitivity by the listeners) in the full-cue condition, and from 0.822 to 0.986 in the edited condition. There was less variability among the ten native English speakers than among the ten subjects in the nonnative groups. For stops presented in the full-cue condition, only six (15%) of the nonnative subjects had A' scores falling within the range of values observed for the native English group. Three of these nonnative subjects were experienced native Spanish speakers of English, two were inexperienced Mandarin speakers of English, and one was an inexperienced Spanish subject. Six (15%) nonnative subjects also had scores falling within the native English range for stops in the edited condition: three inexperienced Spanish subjects, two experienced Spanish subjects, and one inexperienced Mandarin subject. One might conclude from this that at least some of the adult L2 learners were able to master the English word-final /t/-/d/ contrast.

Despite the success of certain individuals, many nonnative subjects—including the so-called "experienced" subjects—produced final stops that were very difficult for the native English listeners to identify correctly. The native speakers' stops were correctly identified in 91% of instances in the two conditions whereas the nonnative subjects' stops were identified in only 64% of instances on the average. One might wonder, therefore, if the nonnatives' stops were produced at above chance rates. Twenty binomial probability tests were carried out to answer this question. An alpha rate of 0.002 for each test resulted in an experiment-wise error rate of 0.10. For stops in the full-cue condition, all four minimal pairs produced by the native English speakers were identified at rates that significantly exceeded the guessing rate of 50% (i.e., correct identifications in more than 123 of 200 instances, p < 0.002). The same was true for all four minimal pairs produced by both the experienced and inexperienced Spanish subjects. Above-chance rates were observed for only three of four minimal pairs produced by the inexperienced Mandarin subjects, and just two minimal pairs produced by the experienced Mandarin subjects.

The final stops in all four minimal pairs produced by the native English speakers were identified at an above-chance rate in the edited condition (p < 0.002), confirming that there were sufficient acoustic cues to the word-final /t/-/d/ distinction in the portion of the CVC waveforms that preceded the stop closure interval (that is, in the preceding vowel and formant transitions; see Flege, 1989). Above-chance rates were obtained for three minimal pairs produced by the experienced Spanish subjects, just one minimal pair produced by the inexperienced Spanish and Mandarin subjects, and no minimal pairs produced by the experienced Mandarin subjects. It appears, then, that three of the four nonnative groups did not produce sufficient preconsonantal cues to the /t/-/d/ distinction.

B. Acoustic analyses

The listening test results failed to support the predictions generated by the SLM (see the Introduction). Both experienced L2 groups differed from the native English speakers, and the final stops produced by the experienced L2 learners were identified no better by the native English lis-
teners than those produced by relatively inexperienced L2 learners. L1 acquisition research has shown that children sometimes produce statistically reliable phonetic differences that go unnoticed by listeners (Maxwell, 1981; Weismer et al., 1981). The purpose of the acoustical analyses presented in this section was to test for between-group differences that might not have been evident to the listeners.

1. Vowel duration

Averaged over the words in all eight minimal pairs, vowels produced by the native speakers of English were longer than those produced by the experienced and inexperienced Spanish speakers (224 vs 180, 154 ms) as well as the experienced and inexperienced Mandarin speakers (169, 154 ms). As shown in Fig. 4, the difference between the native and nonnative speakers was due largely to the duration of vowels preceding /d/. The native English speakers made vowels 87 ms (48%) longer on average before /d/ than /t/. Smaller voicing effects were observed for the experienced Spanish subjects (55 ms or 36%), the experienced Mandarin subjects (37 ms or 25%), the inexperienced Spanish subjects (30 ms or 22%), and the inexperienced Mandarin subjects (32 ms or 23%).

The mean durations of vowels in the /bVt/, /bVd/, /sVt/, and /sVd/ words were submitted to a group X initial consonant (/b/ vs /s/) X final stop (/t/ vs /d/) ANOVA, which yielded a significant group X final stop interaction $[F(4,45) = 7.62, p < 0.001]$. Tests of simple main effects revealed that the talkers in all five subject groups made vowels significantly longer before /d/ than /t/. The $F$ ratios yielded by these tests ranged from 33.3 to 56.7 ($p < 0.001$). The simple main effect of group was nonsignificant for words ending in /t/ $[F(4,45) = 2.01, p > 0.10]$. It was significant, however, for the words ending in /d/ $[F(4,45) = 7.86, p < 0.001]$. Post-hoc comparisons (Tukey's HSD) revealed that the native English speakers made vowels significantly longer before /d/ than the talkers in all nonnative groups ($p < 0.05$). No other difference, including those between the experienced and inexperienced nonnative subjects, reached significance at the 0.05 level.

The mean duration of vowels produced by each talker in eight /t/-final words was subtracted from the mean duration of vowels in eight /d/-final words. As shown in Fig. 5, a few subjects in each nonnative group produced a mean "voicing effect" that fell within the range of mean values observed for the ten native English speakers. If we define the English "norm" as the range of values falling within one standard deviation of the native English speakers' overall mean (50-123 ms in the present instance), then we could credit 11 (28%) of the nonnative speakers with having mastered the "English" voicing effect. Of these, eight were experienced L2 learners and three were relatively inexperienced L2 learners.

The larger number of experienced than inexperienced L2 learners who mastered the English voicing effect is consistent with the hypothesis that the phonetic contrast between English /t/ and /d/ is learnable. An analysis of the ratio of vowel durations preceding /d/ vs /t/ also supported this hypothesis. It is known that the size of voicing effects produced by native speakers of English vary as a function of overall vowel duration, which may depend on factors such as degree of stress and speaking rate (Crystal and House, 1982, 1988). The use of ratios minimizes the potentially confounding effects of such factors. A ratio was therefore computed for each talker by dividing the mean duration of vowels in eight /d/-final words by the mean duration of vowels in eight /t/-final words. The effect of group was marginally significant in a one-way ANOVA examining the vowel duration ratios $[F(4,45) = 3.53, p = 0.014]$. Post-hoc compari-

![FIG. 4. Mean duration of vowels (ms) in minimally-paired words ending in /d/ and /t/ spoken by subjects in five groups. The brackets enclose ± one standard deviation.](image)

![FIG. 5. The mean difference in the duration of vowels in eight words ending in /d/ and eight words ending in /t/ produced by individual subjects in five groups.](image)
sons (Tukey's HSD) revealed that the native English subjects' ratio (1.52) was not significantly greater than the ratios obtained for the experienced Spanish and Mandarin subjects (1.36, 1.26) \( p > 0.10 \). The native speakers' ratios did differ significantly, however, from the inexperienced Spanish and Mandarin subjects' ratios (1.23, 1.25) \( p < 0.05 \).

2. Closure voicing

The duration of closure voicing produced by each subject in /bVt/, /bVd/, /sVt/, and /sVd/ words was examined in a group x initial consonant x final stop ANOVA. As shown in Fig. 6, there was very little closure voicing in the /t/ 's produced by the subjects in any group. All five groups of talkers produced more closure voicing in /d/ than /t/. The native English speakers produced a /d/ vs /t/ difference that averaged 54 ms. The differences produced by the experienced and inexperienced Spanish subjects averaged 40 and 18 ms, respectively; those produced by the experienced and inexperienced Mandarin subjects averaged just 10 and 11 ms, respectively. The significant group x final stop interaction \( F(4,45) = 6.65, p < 0.001 \) yielded by the ANOVA was explored by tests of simple main effects. All five groups of talkers were found to have produced significantly more closure voicing in /d/ than /t/ \( F \)-ratios ranging from 9.77 to 34.1; \( p < 0.01 \). The effect of group on closure voicing was nonsignificant for words ending in /t/ \( F(4,45) = 1.46; p > 0.10 \), but it was significant for /d/-final words \( F(4,45) = 6.53, p < 0.001 \). Post-hoc tests (Tukey's HSD) revealed that the native English speakers sustained voicing significantly longer in /d/ than the subjects in both inexperienced L2 groups. Although they produced more closure voicing in /d/ than the experienced Mandarin subjects \( p < 0.05 \), the native speakers did not differ significantly from the experienced Spanish subjects \( p > 0.10 \), who produced significantly more closure voicing than the inexperienced Spanish subjects \( p < 0.05 \).

The mean closure voicing difference produced by each of the 50 talkers was calculated. One native speaker of English did not produce a closure voicing distinction between /d/ and /t/, a phenomenon that has been noted occasionally before (e.g., Higgs and Hodson, 1978). Of the 40 nonnative subjects, ten (25%) produced a mean closure voicing difference between /t/ and /d/ that fell within one standard deviation of the native English speakers' mean /t/-/d/ difference \( \text{viz.} 24-85 \text{ ms} \). Of these, five were experienced Spanish subjects, four were inexperienced Spanish subjects, and just two were native speakers of Mandarin.

3. Stop closure duration

Due to the missing values discussed in Sec. I, just two mean stop closure durations were calculated for each subject, one for words ending in /t/ and one for words ending in /d/. As shown in Fig. 7, the subjects in all five groups made /t/ closures longer than /d/ closures. The native speakers of English produced a larger average distinction (39 ms) than the experienced Spanish subjects (18 ms), the inexperienced Spanish subjects (7 ms), the experienced Mandarin subjects (27 ms), and the inexperienced Mandarin subjects (16 ms). The mean stop closure duration values were submitted to a group x final stop (/t/ vs /d/) ANOVA. The ANOVA yielded a significant main effect of stop voicing \( F(1,44) = 22.9, p < 0.001 \), a nonsignificant group main effect \( F(1,44) = 1.27, p > 0.10 \), and a nonsignificant two-way interaction \( F(4,44) = 1.54, p > 0.10 \).

4. F1 offset frequency

The native English subjects were expected to produce words ending in /t/ with a higher F1 offset frequency than
words ending in /d/, at least for the two minimal pairs containing non-high vowels (Wolf, 1978; Hillenbrand et al., 1984, Summers, 1987; see, also, Summers, 1988; Fischer and Ohde, 1990). Figure 8 shows the mean frequency of F1 in three 15-ms segments found at the end of the vowel in the /bVt/ and /bVd/ words. (As used here, the term "vowel" includes the formant transitions leading into /t/ and /d/; see Sec. I.) F1 offset frequency was 68 Hz higher, on the average, in /bVt/ than /bVd/ words produced by the English subjects. F1 averaged just 3 Hz higher in /bVt/ than /bVd/ words for the experienced Spanish subjects, and it was actually slightly lower (by 3 Hz) in the /bVt/ than /bVd/ words spoken by the experienced Mandarin subjects.

The frequencies obtained in the final three segments in four /bVt/ and /bVd/ words produced by talkers in the three groups were submitted to separate vowel X final stop X segment ANOVAs. None of the three-way interactions reached significance [F values ranging from 0.82 to 1.83, p > 0.10], suggesting that the offset frequency differences extended into the preceding vowel (see Summers, 1987, 1988). Frequencies in the three segments in each word were therefore averaged and submitted to a (3) group X (2) final stop X (4) preceding vowel ANOVA. The significant three-way interaction it yielded [F(6,81) = 3.64, p < 0.01] was explored by tests of simple main effects. The /t/ vs /d/ difference proved to be significant in all four minimal pairs produced by the native speakers of English [F values ranging from 11.9 to 20.8, p < 0.01]. F1 offset frequency was significantly higher before /t/ than /d/ in bat–bad (739, 610 Hz), bet–bed (592, 485 Hz) and bit–bid (485, 425 Hz); it was significantly lower before /t/ than /d/ in beat–bead (303, 329 Hz). For the experienced Spanish subjects, the F1 offset frequency differences in bat–bad (694, 587 Hz), bet–bed (493, 479 Hz), bit–bid (383, 377 Hz) and beat–bead (365, 380 Hz) were all nonsignificant [F ratios ranging from 0.05 to 2.67, p > 0.10]. The differences in F1 offset frequency produced by the experienced Mandarin subjects in bat–bad (759, 747 Hz), bet–bed (624, 632 Hz), bit–bid (438, 452 Hz) and beat–bead (322, 326 Hz) were also all nonsignificant [F ratios ranging from 0.09 to 1.36, p > 0.10].

F1 offset frequencies for the /sVt/ and /sVd/ words are also shown in Fig. 8. The group X vowel X final stop interaction missed reaching significance in the analysis of the average frequencies obtained in the final 5 ms of each word [F(6,81) = 2.24, p = 0.05]. The significant group X final stop interaction that was obtained [F(2,27) = 14.0, p < 0.001] was explored by tests of simple main effects. The F1 offset frequency difference between /t/ and /d/ produced by the native speakers of English (534 vs 461) proved to be significant [F(1,9) = 76.7, p < 0.01]. The difference produced by experienced Mandarin subjects (543 vs 518 Hz) was also significant [F(1,9) = 11.6, p < 0.01]. The F1 difference produced by the experienced Spanish subjects (471 vs 453 Hz), on the other hand, was nonsignificant [F(1,9) = 4.72, p = 0.058]. A consideration of the frequency values presented earlier reveals that the size of the F1 offset frequency differences in /sVt/ and /sVd/ words was inversely related to vowel height. Not surprisingly, a significant vowel X final stop interaction was obtained in the analysis of both /bVC/ and /sVt/ words [F(3,81) = 6.56, F(3,81) = 13.3; p < 0.001]. Averaged across the three groups, the F1 offset frequency differences decreased steadily from bat–bad (129 Hz) to bet–bed (108 Hz) to bit–bid (59 Hz) to beat–bead (27 Hz); and they decreased steadily from sat–sad (142 Hz) to set–said (116 Hz) to sit–sid (50 Hz) to seat–seed (167 Hz). This finding agrees with previous studies of F1 offset frequency (e.g., Hillenbrand et al., 1984).

5. Discussion

As expected, the native speakers of English made vowels longer before /d/ than /t/, produced more closure voicing in /d/ than /t/, produced words ending in /t/ with a higher F1 offset frequency than /d/-final words, and sustained /t/ closures longer than /d/ closures. The nonnative speakers produced these acoustic distinctions also, but their acoustic differences were usually smaller than the native English speakers'. One exception to this was for stop closure duration. The lack of a significant group X voicing interaction suggested that the nonnative speakers produced as large a closure duration difference between /t/ and /d/ as the native speakers. Also, the experienced Spanish subjects did not differ from the native English speakers in production of closure voicing in /d/. Finally, when the ratios of vowels in words ending in /d/ and /t/ were examined, it was found that the two experienced L2 groups did not differ significantly from the native speakers of English.

Some of the nonnative subjects produced more English-like acoustic distinctions between /t/ and /d/ than others.
Of the 40 who were included in the present study, 73% produced a vowel duration difference in words ending in /t/ vs /d/ that fell within one standard deviation of the mean difference observed for the native English speakers, and 26% did so for the closure voicing difference between /t/ and /d/. To provide a similar estimate for F1 offset frequency, we computed the average F1 frequency in the final 45 ms of the vowels in eight words (beat, bead, bit, bid, bet, bed, bat, bad) produced by the 20 inexperienced nonnative speakers. (Recall that the original F1 analysis examined only experienced nonnative speakers.) The average /t/-/d/ difference in /VC/ words produced by all of the nonnatives was then computed. Of the 40 nonnative subjects, 36% produced F1 /bVC/ words produced by all of the nonnatives was then called that the original F1 analysis examined only experienced nonnative speakers. (Recall that the original F1 analysis examined only experienced nonnative speakers.) The average /t/-/d/ difference in /VC/ words produced by all of the nonnatives was then computed. Of the 40 nonnative subjects, 36% produced F1 frequency differences that fell within one standard deviation of the mean difference produced by the ten native English subjects.

If the percentage of nonnative subjects who achieved the English "norm" is a good indication, one might conclude that it is easier for adult L2 learners to learn the vowel duration correlate of the English /t/-/d/ contrast than the closure voicing or F1 frequency correlates. The former involves the timing of lingual gestures; the latter two involve the coordination of glottal and supraglottal gestures, including gestures that actively enlarge the oral cavity (Flege et al., 1988). Previous research also suggests that children learning English as an L1 succeed in producing the vowel duration cue to final stop voicing before learning to produce other cues such as closure voicing (e.g., Naeser, 1970; Higgs and Hodson, 1978; Smith, 1979; Greenlee, 1980; Weismer et al., 1981; Krause, 1982a,b). One finding obtained in the present study suggested indirectly that phonetic learning may be influenced by lexical frequency. Generally speaking, we found that the nonnative subjects differed more from the native English speakers in producing /d/ than /t/. For example, they closely resembled the native English speakers in producing /t/ with little or no closure voicing, whereas they typically produced /d/ with less closure voicing than the native speakers of English. The nonnative speakers' vowel durations more closely resembled the durations of vowels produced by the native English speakers in the context of /A/ than /d/, something that has been observed in previous L2 production studies (see Flege, 1988a, for a review). These two findings agree with the linguistic description of /d/ as more "marked" in word-final position than /t/. It is probably the result of differences in ease of articulation that more human languages permit thousands of times before they can accurately produce a new sound in it.

C. Multiple correlation analyses

Forward stepwise multiple regression analyses were carried out to examine the relationship between the A' scores obtained for minimal pairs produced by the nonnative speakers (Sec. II A) and acoustic measures of those minimal pairs (Sec. II B). Three acoustic predictor variables were regressed onto the 160 A' scores obtained for the nonnative speakers: differences in the duration of vowels in the two members of each minimal pair; closure voicing differences; and the F1 offset frequency differences. (Stop closure duration values were not entered into the model because of missing data; see above.) Results of the analyses of A' scores obtained in the edited and full-cue conditions are summarized in Table III.

Variations in vowel duration accounted for 15.2% of the variance in the full-cue condition at step 1. Closure voicing accounted for an additional 3.4% of the variance at step 2; and F1 offset frequency accounted for an additional 1.5% of the variance at step 3. The seemingly greater importance of vowel duration than closure voicing or F1 offset frequency was not an artifact of the order in which the variables were entered into the forward stepwise regression model. When voicing was forced as the first variable, it accounted for only 6% of the variance. Vowel duration accounted for 9% more variance at step 2. When F1 offset frequency was forced as the first variable, it accounted for 4% of the variance, with vowel duration accounting for an additional 14% of the variance at step 2.

It would probably be unwise to assume that the order in which the three variables were entered into the regression model, or their beta weights, indicates the variables' relative perceptual importance to native English-speaking listeners. The order of entry may simply reflect the size of the acoustic contrasts produced by the nonnative speakers. Recall that

| TABLE III. Summary of analyses examining acoustic measures of four minimal pairs spoken by 40 nonnative speakers of English and listeners' perceptual sensitivity (A') to the /t/-/d/ contrast in these minimal pairs in two listening conditions. The adjusted R^2 values were obtained at steps 1–3 of forward stepwise multiple regression analyses. |
|-----------------------------|-------------------|----------------|-------------|
|                            | R^2               | Beta           | F           | p            |
|                            | Full-cue condition |                |             |             |
| VOWDIFF^a                  | 0.152             | 0.343          | 22.28       | 0.000**      |
| VOIDIFF^b                  | 0.186             | 0.183          | 6.36        | 0.031*       |
| F1DIFF^c                   | 0.210             | 0.155          | 4.67        | 0.032*       |
|                            | Edited condition  |                |             |             |
| VOWDIFF^d                  | 0.280             | 0.501          | 53.62       | 0.000*       |
| VOIDIFF^e                  | 0.300             | 0.087          | 1.62        | 0.205        |
| F1DIFF^f                   | 0.293             | 0.120          | 3.18        | 0.076        |

^a VOWDIFF—the difference in the duration of vowels preceding /t/ vs /d/ in each of four minimal pairs produced by the 40 nonnative talkers.
^b VOIDIFF—the difference in the duration of closure voicing in /t/ vs /d/.
^c F1DIFF—the difference in offset frequency of F1 in the 45-ms interval preceding constriction of word-final /t/’s and /d/’s.

Flege et al.: Production of the word-final /t/-/d/ contrast

73% of the nonnative speakers produced vowel duration differences that fell within one standard deviation of the mean difference produced by the native English speakers, whereas this was true for only 26% of the nonnative speakers for closure voicing, and 36% of the nonnative speakers for F1 offset frequency.

We can think of two explanations for why the regression model accounted for only 21% of the variance in A’ scores obtained in the full-cue condition. First, acoustic measures of the spectrum and intensity of final release bursts were not included in the model. Second, the nonnative speakers simply did not produce very large acoustic distinctions between /t/ and /d/. For example, the perceptual effect of vowel duration may have accounted for relatively little of the variance because the stop voicing effects produced by the nonnative speakers was far smaller than the effects produced by the native English speakers (37 vs 83 ms on the average). In fact, when the acoustic data for the ten native English subjects were included in a new regression analysis, the variance that was accounted for (largely by vowel duration) increased from 21% to 37%. Another factor that must be considered is that vowel duration is a relative rather than an absolute cue. The listeners heard five native and 20 nonnative speakers in each block. The voicing effect produced by a nonnative speaker with a characteristically fast speaking rate might have gone unnoticed if the vowels he/she produced in the context of /d/ were shorter than those produced in the context of /t/ by subjects with a characteristically slowly speaking rate. Perhaps vowel duration would have accounted for more variance had we presented the CVCs in their original carrier phrases.

We hypothesized that vowel duration would account for more variance in the edited than in the full-cue condition. This hypothesis, which will be referred to as the “attentional allocation hypothesis,” supposed that the listeners would allocate greater attention to vowel duration in the edited condition because other cues to the /t/-/d/ distinction, such as closure voicing and release bursts cues, would be unavailable in the edited condition. In fact, vowel duration accounted for more variance when it was entered at step 1 in a stepwise multiple regression analysis examining A’ scores obtained in the edited condition than it did in the analysis of full-cue condition data (28% vs 15%). Neither closure voicing nor F1 offset frequency were identified as significant predictors of edited-condition A’ scores.

To provide a further test of the attentional allocation hypothesis, we tested the strength of the correlation between vowel duration differences and the A’ scores that were obtained for the minimal pairs in both the edited and the unedited conditions. One variable in the two correlational analyses (i.e., vowel duration differences in the 160 minimal pairs) was the same, so we carried out a test of the equality of two related correlations (Kleinbaum et al., 1988). As predicted, the correlation involving A’ scores obtained in the edited condition was significantly stronger than the correlation involving A’ scores obtained in the full-cue condition (Z = 2.54, p < 0.01). Recall that the same vowel duration differences were available to the listeners in both conditions. Only cues that occurred after constriction of the final stop were removed in the edited condition. This finding therefore suggests that the listeners made greater use of the small, albeit significant, vowel duration differences produced by the nonnative speakers in the edited condition than they did in the full-cue condition.

The finding just presented is very striking when one considers that some of the nonnative speakers produced very small vowel duration differences between words ending in /t/ vs /d/. We reasoned that the difference in the strength of correlations obtained in the two listening conditions would have been even greater had we chosen to examine only minimal pairs with relatively large vowel duration differences. To test this, we performed separate analyses of the 80 minimal pairs having the highest A’ scores in the full-cue condition and the 80 pairs with the lowest A’ scores. These might be called the relatively “intelligible” and “unintelligible” minimal pairs. Vowel duration differences were significantly greater in the intelligible than in the relatively unintelligible minimal pairs (52 vs 22 ms) [F(1,158) = 18.5, p < 0.01]. There was not a significant difference in the size of the F1 offset frequency differences in the intelligible and unintelligible minimal pairs (15 Hz vs −1 Hz) [F(1,158) = 2.00, p > 0.10]. Closure voicing differences between /t/ and /d/ were significantly larger for the intelligible than the unintelligible minimal pairs (26 ms vs 12 ms) [F(1,158) = 6.90, p < 0.01] but, of course, this would not have been evident to listeners in the edited condition.

We computed the simple correlations between vowel duration differences in the two sets of 80 minimal pairs and the A’ scores obtained for those two sets of pairs in both the edited and full-cue conditions. If the native English listeners allocated more attention to vowel duration in the edited than in the full-cue condition, we might expect to find a significant difference in the strength of correlations in the first but not the second analysis because much larger vowel duration differences were available in the intelligible than in the unintelligible minimal pairs. In fact, the correlation obtained for the intelligible minimal pairs was significantly stronger in the edited than full-cue condition [r = 0.600 vs r = 0.349; Z = 2.43, p < 0.01] but there was not a significant difference in strength of correlation for A’ scores obtained for the unintelligible minimal pairs in the edited and full-cue conditions [r = 0.218 vs r = 0.078; Z = 1.12, p > 0.10]. This implies that a vowel duration difference that averages just 22 ms cannot be exploited perceptually by native speakers of English whereas a difference averaging 52 ms can be used.

III. GENERAL DISCUSSION

A. L2 learners’ stop production

Previous studies have shown that adults who learn English as a second language in adulthood often do not produce a fully English-like distinction between voiced and voiceless stops in the final position of English words (e.g., Suomi, 1976; Elsendoorn, 1980, 1982; Flege and Port, 1981; Eckman, 1981; Mitleb, 1981a,b; Port and Mitleb, 1983; Anderson, 1983, 1987; Heyer, 1986; Weinberger, 1987; Lowie, 1988; Flege and Davidian, 1984; Flege et al., 1988; Flege,
Based on the nonoccurrence of word-final stop consonants in Mandarin and Spanish (see the Introduction), we assumed that native speakers of those languages would not identify word-final English /t/ and /d/ tokens in terms of word-final L1 phonetic categories. Given this assumption, the speech learning model under evaluation (e.g., Flege, 1988a, 1992a,b) generated the prediction that adult native speakers of Spanish and Mandarin, at least those who were experienced in English, would produce /t/ and /d/ if they identified both English /b,d,g/ and /p t k/ in terms of their L1 /p t k/ categories.

Acoustic analyses revealed that the native speakers distinguished /t/ from /d/ on the basis of vowel duration, F1 offset frequency, the presence/absence of closure voicing, and stop closure duration. The nonnative speakers produced the same acoustic differences, but their differences were generally smaller in magnitude than the native English speakers'. This might account for why the native English speakers' production of the word-final /t/-/d/ contrast was consistently smaller in magnitude than the native English speakers' production of the word-final /t/-/d/ contrast in previous studies, this might be attributed to the lack of sufficient L2 experience. For others, it might be attributed to the interlingual identification of L1 and L2 stops. For example, one might expect French learners of English to show a smaller effect of stop voicing on the duration of preceding vowels than native speakers of English if they identified word-final tokens of English /b d g/ and /p t k/ in terms of the corresponding stop consonants of French. This is because the stop voicing effect is smaller in French than English (Mack, 1982). So too, one might expect adults whose L1 permitted only /p t k/ to occur in word-final position to continue "devoicing" English /b,d,g/ if they identified both English /b,d,g/ and /p t k/ in terms of their L1 /p t k/ categories.

Listeners' perceptions of the voicing contrast in word-final position, however, this was not always true. There, the most common misidentification of stops produced by both the native and nonnative speakers was /t/ for /d/.

Second, we may have underestimated the nonnatives' ability to produce word-final stops if we did not observe their optimal L2 speech production. Our subjects read isolated CVC words from a list. It was assumed that this would optimize their pronunciation of English by allowing them to "monitor" their speech. This assumption should not be allowed to go unchallenged, however. There is evidence, for example, that nonnatives' spontaneous conversations may sound less foreign accented than their reading of paragraph-length texts (Oyama, 1976).

Third, the study may have yielded a negative finding because even the experienced L2 learners had not received sufficient phonetic input to master the English /t/–/d/ contrast. The experienced L2 learners had, of course, lived far longer in the US than the inexperienced subjects (7.3 vs 0.7 years on the average). However, we believe that L1 acquisition must be considered in attempting to define what constitutes "sufficient" L2 input. Most English-learning children produce a perceptually effective contrast between /t/ and /d/ in word-final position by the age of five years (e.g., Naeser, 1970; Higgs and Hodson, 1978; Smith, 1979; Greenlee, 1980; Weismer et al., 1981; Krause, 1982a, b; Smit et al., 1990). One might therefore suppose that five or more years of residence in the US would surely have provided the opportunity for enough phonetic input for mastery of the /t/–/d/ contrast if adults, as claimed, are able to master new L2 consonants.

There is no guarantee, however, that our subjects did receive sufficient phonetic input. Consider, for example, the nonnative subjects' report of using English on average 57% of the time on a daily basis. Were their self-reports accurate? If so, would an adult who had lived for 8 years in the US but spoke English 57% of the time have received as much English input as the average 5-year-old English monolingual child? And would the English input received by such an adult be as useful as the input received by young children learning English as an L1? It may be that young children learning an L1 hear more isolated words than adults learning an L2, and that this may be helpful in learning to pronounce an L2. For example, Brown (1973) suggested that young children are aided in learning to identify word units in the model's predictions were not confirmed by the present study, although the present data will not permit us to choose among the alternate explanations offered. First, we may have underestimated the nonnatives' ability to produce word-final stops in English if alveolar stops are more difficult to learn than bilabial and velar stops. English /t/ and /d/ are variable in production, sometimes being realized as flaps (intervocalic position) or glottal stops (word-final position). Takata and Nabelek (1990) examined the effect of noise and reverberation on the identification of English CVC words spoken by native speakers of English and Japanese. Listeners' confusions could be readily understood on the basis of cross-language phonological differences for consonants found in word-initial position (e.g., the Japanese subjects' confusion of /r/ and /l/). In word-final position, however, this was not always true. There, the most common misidentification of stops produced by both the native and nonnative speakers was /t/ for /d/.
speech stream by the fact that their parents often speak in isolated words to them. Vogel and Winitz (1989) provided evidence that having been exposed to isolated Chinese words helped native English adults identify the boundaries between words in Chinese sentences.

Consider also the effect of stop voicing on preceding vowel duration. It is generally agreed that the relatively large voicing effect seen in English is an exaggerated version of an effect seen in other languages. Vowel duration is just one of several cues to the voicing distinction in word-final English stops. Perhaps because of this, the English voicing effect is not always maintained. It is far smaller in connected speech than in isolated words, especially in words not found in utterance-final position or in pre-pausal positions within utterances (e.g., Luce and Charles-Luce, 1985; Crystal and House, 1988). Children learning English as an L1 may receive more useful phonetic input than adults learning an L2 as the result of an enhancement of the vowel duration correlate of final stop voicing. Bernstein Ratner (1986) observed a greater magnitude of prepausal lengthening in the speech mothers addressed to children just beginning to talk than in the speech those mothers addressed to other adults. Such an increase was less evident in the speech mothers addressed to children who were at a slightly more advanced stage of language development. One might suppose that the voicing effect would be augmented in words whose overall durations have been increased by prepausal lengthening. In fact, Bernstein Ratner and Luboff (1984) reported that the voicing effect was larger in the speech mothers addressed to children just learning to talk than in speech directed to other adults (roughly 80 vs 40 ms). We suspect that adults do not exaggerate voicing effects when speaking to nonnative adults, perhaps as an indirect result of producing fewer words in prepausal positions owing to the use of longer, more complicated sentences.

Still another factor that must be considered is the role of foreign-accented input. We suspect that our subjects spoke English with at least some nonnative speakers who failed to distinguish /t/ from /d/ in word-final position. Perhaps our nonnative subjects would have produced better /t/-/d/ contrasts had they received better (i.e., more authentic) English-language speech input. Some of the nonnative subjects may have failed to produce native-like /t/-/d/ contrasts because they established phonetic categories defined both by native and foreign-accented /t/ and /d/ tokens. An effect of foreign-accented input could be ruled out if results comparable to those obtained here were obtained from experienced late L2 learners who had little opportunity to speak their L1 and no opportunity to speak English with other nonnatives.

We must, of course, face the possibility that the predicted effects were not observed because the model is incorrect in one or more respects. For example, the SLM may be wrong in claiming that L2 production difficulties have a perceptual origin. The L2 learners examined here may not have identified word-final English /t/ and /d/ tokens in terms of existing L1 categories, thereby blocking the establishment of phonetic categories for word-final English /t/ and /d/. They may have established central perceptual representations for English /t/ and /d/ but were unable to motorically output the sounds specified in those representations. If so, the data at hand would support the hypothesis of a critical period for human speech learning (Scovel, 1988; Patkowski, 1989) and could be interpreted to mean that adults are unable to master L2 sounds not present in their L1 phonetic inventory.

Alternatively, the SLM may be correct in claiming that production difficulties have a perceptual origin but incorrect in positing that interlingual identification occurs at a phonetic level. That is, contrary to the model, the experienced L2 learners may have identified word-final /t/ and /d/ tokens in terms of phonetic categories used for word-initial /t/ and /d/ in the L1. Direct support of this was provided recently by Flege (1989; see, also, Flege and Wang, 1990). If it could be shown directly that word-final L2 stops are identified with word-initial L1 stops despite substantial acoustic differences between word-initial and word-final stops, it would support the hypothesis of Trubetzkoy (1939) that L2 learners "filter out" acoustic differences between corresponding L1 and L2 sounds that are not phonemically relevant for contrasts in the L1. (See Jaeger, 1980, 1986, for experiments pertaining to listeners' perceptual grouping of word-initial and word-final allophones into phonemic units.) Such evidence, should it be forthcoming, would suggest that the interlingual identification of L1 and L2 sounds occurs at a phonemic level rather than at the phonetic level stipulated by the SLM. If so, one would expect native Mandarin and Spanish learners of English to be unable to establish phonetic categories for word-final English /p/, /t/, and /k/ and, accordingly, to produce these stops inaccurately in English. Additional research is needed to determine if Spanish and Mandarin learners of English to identify word-final English stops in terms of word-initial L1 stop categories, and if they (but not native English speakers) use the same acoustic features in identifying the distinction between /t/ and /d/ in both the word-initial and word-final positions.

Finally the SLM may be wrong in claiming that all adult learners who have received sufficient L2 phonetic input will master new consonants in an L2. Perhaps new consonants can be mastered by only a small proportion of adult L2 learners. Our subjects were not selected on the basis of a special ability to pronounce English. The first 40 adult L2 learners who met very broad selection requirements (see Sec. I) were enrolled in the study. Of these, 15% produced a perceptually effective /t/-/d/ contrast (that is, showed A' values falling within one standard deviation of the mean A' value obtained for ten native English speakers). In an average of 45% of instances, the nonnatives produced acoustic differences between /t/ and /d/ (in vowel duration, duration of closure voicing, F1 offset frequency) that fell within one standard deviation of the mean differences produced by the native English subjects. Although it is clear that not all subjects mastered the English /t/-/d/ contrast, it seems reasonable to conclude that at least some of them did so.

B. Listeners' perception of foreign-accented stops

As noted earlier, the nonnative speakers produced a much smaller difference in the duration of vowels preceding /d/ vs /t/ than the native speakers of English (37 vs 83 ms).
Despite this, a multiple regression analysis revealed that vowel duration accounted for a small but significant amount of variance in the listeners' identification of the nonnative speakers' stops. We hypothesized that the native English-speaking listeners would make greater use of vowel duration differences when CVCs produced by the nonnative speakers were presented in an "edited" condition (where closure voicing and release burst cues were absent) than in the original "full-cue" condition. This hypothesis was motivated by three considerations. First, a study by Flege and Hillenbrand (1987) suggested that native speakers of a language in which word-final stops are consistently released (viz., French) make less use of closure voicing and F1 offset cues than native speakers of a language in which final stops are not consistently released (viz., English; see, e.g., Crystal and House, 1982; Bernstein Ratner, 1984). Results of that study suggested that listeners may tailor their perceptual processing of stops to the range of cues that are reliably available. Second, listeners seem to reorganize perceptual processes when adjusting to foreign-accented speech. After a period of acclimation, the foreign-accented speech becomes more intelligible (Wingstedt and Schulman, 1984). Finally, speech perception research has shown that listeners' ability to attend selectively to portions of a word is very flexible and precise (see Pitt and Samuel, 1990, and references therein).

The attentional allocation hypothesis was supported. We noted a significantly stronger correlation between the size of vowel duration differences and A' scores obtained in the edited condition than between the same vowel duration differences and A' scores obtained in the full-cue condition. This suggests that the listeners allocated greater attention to preconsonantal cues such as vowel duration and F1 offset frequency when information pertaining to the constriction and release of final stops was not available in a listening condition. Such flexibility in the face of a varying mix of acoustic cues to a phonetic contrast is just what one would expect given the great variety in speech that arises from differences in dialect, foreign accents, and speaking styles.

The conclusion concerning attentional allocation must be tempered somewhat for two reasons. First, listeners in the present study all participated in the edited condition after participating in the full-cue condition. Perhaps they made greater use of vowel duration in the second condition because they had grown better able to detect small vowel duration differences as the result of their experience identifying stops in the full-cue condition. Second, we did not directly observe a shift in attentional allocation. A more direct test of the hypothesis might be carried out by comparing the performance of native speakers of English and French. English /t/ and /d/ are produced with little or no closure voicing at the end of English words (Smith, 1979; Hogan and Rozsypal, 1980; Guy, 1980; Crystal and House, 1982; Bernstein Ratner, 1984). Perhaps as the result of such variability, the removal of closure voicing and final bursts seldom affected greatly native English listeners' identification of the voicing feature in stops (e.g., Wolf, 1978; Revoile et al., 1982; Wardrip-Fruin, 1982; Hillenbrand et al., 1984; Flege, 1989; but see Malécot, 1958). In French, on the other hand, word-final stops are usually released, and /b d g/ are often produced with voicing through the entire period of constriction (see Flege and Hillenbrand, 1987). As noted earlier, the effect of stop voicing on preceding vowel duration appears to be greater in English than French (Mack, 1982). One might therefore expect that native English subjects, as the result of their experience with a highly variable /t/-/d/ contrast, would be more flexible than French speakers to varying mixes of stop voicing cues. More specifically, one might expect to observe a greater difference in the strength of correlations between vowel duration and perceptual judgments in edited versus full-cue conditions for native English than native French subjects.

In summary, the present study failed to confirm the prediction (Flege, 1992a,b) that experienced native Mandarin and Spanish speakers of English would master the production of /t/ and /d/ in the final position of English words. Several alternative explanations for this negative finding were offered, but the data at hand did not permit choosing among them. Native English-speaking listeners were shown to make perceptual use of the small but highly significant vowel duration produced by the nonnative speakers. A stronger correlation was noted between vowel duration differences present in minimally paired words ending in /d/ vs /t/ and A' scores obtained for the identification of stops in those words when the words were presented in an edited than in a full-cue listening condition. This was interpreted to mean that the listeners made greater perceptual use of vowel duration when acoustic cues pertaining to constriction and release were systematically absent than when those cues were present in the signal.

ACKNOWLEDGMENTS

This research was funded by NIH grant DC00257. The authors thank C. Wang for help with data collection; C. Crowther, V. Mann, R. Port, A. Walley, and an anonymous reviewer for comments on a previous draft of the article; and M. Beckman and J. Maddieson for providing cross-language phonetic information.

\footnote{Two Chinese subjects in the study were native speakers of Mandarin, which permits no word-final stops. The remaining subject was a native speaker of Hakka (Keja), a language which, like Taiwanese, permits /ptk/ in word-final position.}


